

## For Whom the Gate Tolls?

### How and Why to Free the Refereed Research Literature Online Through Author/Institution Self-Archiving, Now

*Stevan Harnad* \*

**Abstract:** All refereed journals will soon be available online; most of them already are. This means that anyone will be able to access them from any networked desk-top. The literature will all be interconnected by citation, author, and keyword/subject links, allowing for unheard-of power and ease of access and navigability. Successive drafts of pre-refereeing preprints will be linked to the official refereed draft, as well as to any subsequent corrections, revisions, updates, comments, responses, and underlying empirical databases, all enhancing the self-correctiveness, interactivity and productivity of scholarly and scientific research and communication in remarkable new ways. New scientometric indicators of digital impact are also emerging (<http://opcit.eprints.org>).

to chart the online course of knowledge. But there is still one last frontier to cross before science reaches the optimal and the inevitable: Just as there is no longer any need for research or researchers to be constrained by the access-blocking restrictions of paper distribution, there is no longer any need to be constrained by the impact-blocking financial fire-walls of Subscription/Site-License/Pay-Per-View (S/L/P) tolls for this give-away literature. Its authors/researchers have always donated their research reports for free (and its

---

\* Address all communications to: Stevan Harnad, Intelligence/Agents/Multimedia Group, Department of Electronics and Computer Science, University of Southampton, Highfield, Southampton, SO17 1BJ United Kingdom; E-Mail: [harnad@ecs.soton.ac.uk](mailto:harnad@ecs.soton.ac.uk), URL: <http://www.cogsci.soton.ac.uk/~harnad/>

Es handelt sich hier um eine leicht überarbeitete Fassung des Beitrages: Harnad, Stevan (2003). Open Access to Peer-Reviewed Research Through Author/Institution Self-Archiving: Maximizing Research Impact by Maximizing Online Access. In: Law, Derek & Judith Andrews, Eds. *Digital Libraries: Policy Planning and Practice*. Ashgate Publishing 2003. <http://www.ecs.soton.ac.uk/~harnad/Temp/digital-libraries.htm>

referees/researchers have refereed for free), with the sole goal of maximizing their impact on subsequent research (by accessing the eyes and minds of fellow-researchers, present and future) and hence on society. Generic (OAI-compliant) software is now available free so that institutions can immediately create Eprint Archives in which their authors can self-archive all their refereed (published) papers for free for all forever (<http://www.eprints.org/>). These interoperable Open Archives (<http://www.openarchives.org>) will then be harvested into global, jointly searchable "virtual archives" (e.g., <http://arc.cs.odu.edu/>). "Scholarly Skywriting" in this PostGutenberg Galaxy will be dramatically (and measurably) more interactive and productive, spawning its own new digital metrics of productivity and impact, allowing for an online "embryology of knowledge."

## An Anomalous Picture

*What is wrong with this Picture?*

- 1) A brand-new PhD recipient proudly tells his mother he has just published his first article. She asks him how much he was paid for it. He makes a face and tells her "nothing," and then begins a long, complicated explanation ...
- 2) A fellow-researcher at that same university sees a reference to that same article. He goes to their library to get it: "It's not subscribed to here. We can't afford that journal. (Our subscription/license/loan/copy budget is already overspent.)"
- 3) An undergraduate at that same university sees the same article cited on the Web. He clicks on it. The publisher's website demands a password: "Access Denied: Only pre-paid subscribing/licensed institutions have access to this journal."
- 4) The undergraduate loses patience, gets bored, and clicks on *Napster* to grab an MP3 file of his favorite bootleg CD to console him in his sorrows.
- 5) Years later, the same PhD is being considered for tenure. His publications are good, but they're not cited enough; they have not made enough of a "research impact." Tenure denied.
- 6) Same thing happens when he tries to get a research grant: His research findings have not had enough of an impact: Not enough researchers have read, built upon and cited them. Funding denied.

- 7) He decides to write a book instead. Book publishers decline to publish it: "It wouldn't sell enough copies because not enough universities have enough money to pay for it. (Their purchasing budgets are tied up paying for their inflating annual journal subscription/license/loan costs ...)"
- 8) He tries to put his articles up on the Web, free for all, to increase their impact. His publisher threatens to sue him and his server-provider for violation of copyright.
- 9) He asks his publisher: "Who is this copyright intended to protect?" His publisher replies: "You!"

*What is wrong with this picture?*

(And why is the mother of the PhD whose *give-away* work people cannot steal, even though he wants them to, in the same boat as the mother of the recording artist whose *non-give-away* work they can and do steal, even though he does not want them to?)

## Resolving the Anomaly:

How a few critical *distinctions* plus a few simple *actions* can restore sense to it all.

### 1. Five Essential PostGutenberg Distinctions:

In order to understand what is wrong with the picture, you first have to make *five critical distinctions*. If you fail to make any one of these distinctions, it will be impossible to make sense of the picture or to resolve the anomaly, an anomaly completely unique to the online era of "Scholarly Skywriting" (Harnad 1990) in the "PostGutenberg Galaxy" (Harnad 1991).

#### 1.1 Distinguish the *non-give-away* literature from the *give-away* literature

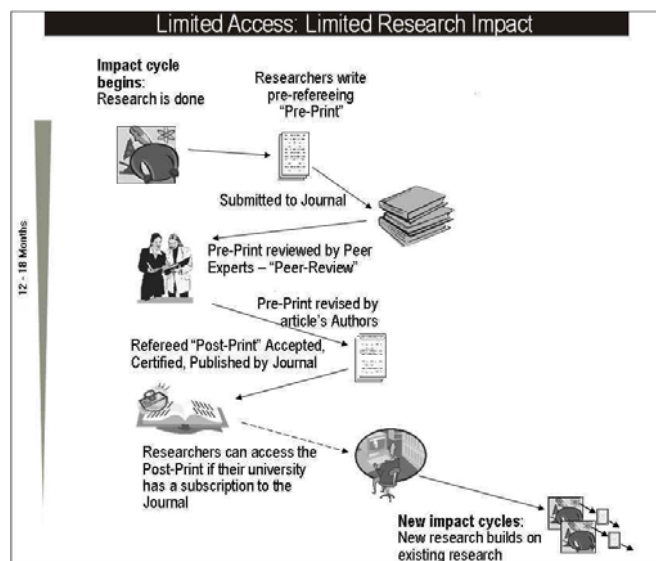
This is the most important PostGutenberg distinction of all. It is what makes this small refereed research literature anomalous (about 24,000 refereed journals, about 2,500,000 articles annually) – fundamentally unlike the bulk of the written literature: Its authors do not seek, nor do they receive, royalties or fees for their writings. Their texts are author give-aways (Harnad 1995a). The only thing these authors seek is research "impact" (Harnad & Carr 2000), which comes from accessing the eyes and minds of all potentially interested fellow-researchers everywhere, now, and any time in the future.

The litmus test for whether a piece of writing falls in the small give-away sector of the literature or the much larger non-give-away sector is: "Does the author seek a royalty or fee in exchange for his writings?" If the answer is *yes* (as it is for virtually all books [cf. Harnad, Varian & Parks 2000] and newspaper or magazine articles), then the writing is *non-give-away*; if the answer is *no*, then it is give-away.

*None* of what follows here is applicable to *non-give-away* writing, but the non-give-away model is the one that most people have in mind for all of writing. So it is not surprising that that small fraction of writing that the more general model does *not* fit should seem anomalous.

## 1.2 Distinguish *income* (arising from article *sales*) from *impact* (arising from article *use*)

Unlike all other authors, researchers derive their income not from the *sale* of their research reports but from the scholarly/scientific *impact* of their reported findings, i.e., how much they are read, cited, and built-upon by other researchers. Hence all fee-based access-barriers are income-barriers for research and researchers (Harnad 1998a), restricting their potential impact to only those (institutions, mainly) who can and do pay the access-fees.



As most institutions cannot afford the access-fees to most refereed research journals, this means that most research papers cannot be accessed by most researchers (Harnad 1998b): Currently, all that potential impact is simply lost.

Note that although researchers do not derive income from the *sale* of their refereed research papers ("imprint income"), they do derive income from the *impact* of those papers ("impact income").

The simple reason why researchers, unlike non-give-away authors, do not seek imprint-income for their refereed research is that the access-tolls for collecting imprint-income are barriers to impact-income (research grants, salaries, promotion, tenure, prizes), which is by far the more important reward for researchers, most of whose refereed papers are so esoteric (Harnad 1995b) as to have no imprint-income market at all.

### 1.3 Distinguish between copyright protection from *theft-of-authorship* (plagiarism) and copyright protection from *theft-of-text* (piracy)

These two very different aspects of copyright protection have always been conflated (Harnad 1999b), because it is the much larger and more representative *non-give-away* literature that has always been the model for copyright law and copyright concerns. But copyright protection from theft-of-authorship (plagiarism), which is essential for both give-away and non-give-away authors, has nothing at all to do with copyright protection from theft-of-text (piracy), which non-give-away authors want but give-away authors do *not* want. One can have full protection from plagiarism without seeking any protection from piracy.

### 1.4 Distinguish *self-publishing* (vanity press) from *self-archiving* (of published, refereed research)

The essential difference between unrefereed research and refereed research is quality-control (peer review, Harnad 1998/2000) and its certification (by an established peer-reviewed journal of known quality). Although researchers have always wished to give away their refereed research findings, they still wish them to be refereed, and certified as having met established quality standards. Hence the *self-archiving* of refereed research should in no way be confused with *self-publishing*, for it includes as its most important component, the online self-archiving, free for all, of refereed, *published* research papers.

### 1.5 Distinguish unrefereed preprints from refereed postprints ("eprints" = preprints + postprints)

Eprint archives, consisting of research papers self-archived online by their authors, are not, and have never been, merely "preprint archives" for unrefereed research. Authors can self-archive therein all the embryological stages of the

research they wish to report, from pre-refereeing, through successive revisions, till the refereed, journal-certified postprint, and thence still further, to any subsequent corrected, revised, or otherwise updated drafts (post-postprints), as well as any commentaries or responses linked to them. These are all just way-stations along the scholarly skywriting continuum.

## 2. The Optimal and Inevitable for Researchers

- The entire full-text refereed corpus online
- On every researcher's desktop, everywhere
- 24 hours a day
- All papers citation-interlinked
- Fully searchable, navigable, retrievable
- For free, for all, forever

All of this will come to pass. The only real question is "How soon?" Will we still be *compos mentis* and fit to benefit from it, or will it only be for the napster generation? Future historians, posterity, and our own still-born scholarly impact are already poised to chide us in hindsight (Harnad 1999b).

What can the research community do to hasten the optimal and inevitable? Here are some recent concepts that may help.

## 3. Two Useful Acronyms, One New Distinction, and One New Ally

### 3.1 S/L/P [Subscription/Site-License/Pay-Per-View] tolls: The impact/access-barriers

Access-tolls are the access-barriers, hence the impact-barriers, for researchers and their give-away research. Access-tolls are the journal publisher's means of recovering costs and making a fair profit. High costs were inescapable in the expensive and inefficient on-paper Gutenberg era; but today, in the on-line PostGutenberg era, continuing to do it all the old Gutenberg way, with its high costs, must be clearly seen as the optional add-on (for this give-away literature only: not for the royalty/fee-based literature!) that it has become, rather than as the obligatory feature it used to be.

Beware of the language of obligatory "value-added," with which the peer-reviewed literature must, by implication, continue to be inextricably wrapped. The only *essential* service still provided by journal publishers (for this anomalous, author-give-away literature in the PostGutenberg era) is peer review itself.

The rest – on-paper versions, PDF on-line page images, deluxe online enhancements – are all potentially valuable features, to be sure, but only as take-it-or-leave-it options. In the on-line era there is no longer any necessity, hence no longer any justification whatsoever, for continuing to hold the refereed research itself hostage to access-tolls and whatever add-ons they happen to pay for.

Beware also of any attempt to trade off S for L or L for P: Pick your poison, all three are access-barriers, hence impact-barriers, and hence all three must go – or rather, they must all now become only the price-tags for the add-on, deluxe *options* that they buy for the researcher and his institution, but no longer also for the peer-reviewed *essentials*, which can now be self-archived for free for all.

### 3.2 Peer review

Peer review itself is *not* a deluxe add-on for research and researchers: This quality-control service and its certification is an *essential* (Harnad 1998/2000). Without peer review the research literature would be neither reliable nor navigable, its quality uncontrolled, unfiltered, un-sign-posted, unknown, unaccountable.

But the peers who review it for the journals are the researchers themselves, and they review it for free, just as the researchers report it for free. So it must be made quite clear that the only real peer-review cost is that of *implementing* the peer review, not actually *performing* it.

Estimates (e.g., Odlyzko 1998) as well as the real experience of online-only journals (e.g., Journal of High Energy Physics <http://jhep.cern.ch/>; Psycology <http://www.cogsci.soton.ac.uk/psycology/>) have shown that the peer-review implementation cost is quite low – about 10-30% of the total amount that the world's institutional libraries (or rather, the small subset of them that can afford any given journal at all!) are currently paying annually *per article* in access-tolls.

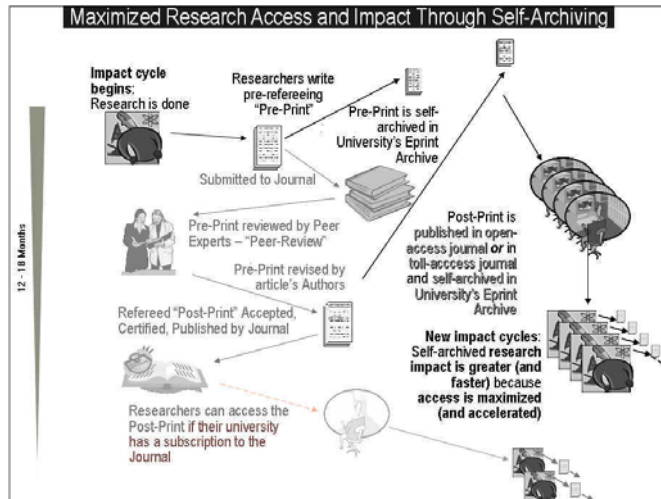
Once the 70-90% toll-based add-ons become optional, the essential 10-30% peer-review cost could easily be paid out of the 100% toll savings – if ever the world's libraries decide they no longer need the add-ons. (The other 70-90% savings can be used to buy other things, e.g., books, which are not, and never will be, author give-aways.)

### 3.3 Separating (i) peer-review *service-provision* from (ii) eprint *access-provision* (and from (iii) optional add-ons)

Researchers need not and should not wait until journal publishers voluntarily decide to separate the provision of the essential peer-review *service* from all the other optional add-on *products* (on-paper version, publisher's PDF version,

deluxe enhancements) before their give-away refereed research can at last be freed of all access- and impact-barriers.

All researchers can free their own refereed research now, virtually overnight, by taking the matter into their own hands; they can self-archive it in their institutional Eprint Archives: <http://www.eprints.org>. Access to the eprints of their refereed research is then immediately freed of all toll-barriers, forever.



### 3.4 Interoperability: The Open Archive Initiative (OAI)

Papers self-archived by their authors in their institutional Eprint Archives can be accessed by anyone, anywhere, with no need to know their actual location, because all Eprints Archives are compliant with the Open Archives Initiative (OAI) meta-data tagging protocol for interoperability: <http://www.openarchives.org>.

Because of their OAI-compliance, the papers in all registered Eprints Archives can be harvested and searched by Open Archive Services such as Cite-Base (<http://cite-base.ecs.soton.ac.uk/help/index.php3>) and the Cross Archive Searching Service (<http://arc.cs.odu.edu/>), providing seamless access to all the eprints, across all the Eprint Archives, as if they were all in one global, virtual archive.



## 4. The Subversive Proposal

### 4.1 Enough to free entire refereed corpus, forever, immediately

Eight steps will be described here. The first four are not hypothetical in any way; they are guaranteed to free the entire refereed research literature (~24K journals annually) from its access/impact-barriers right away. The only thing that researchers and their institutions need to do is to take these first four steps. The second four steps are hypothetical predictions, but nothing hinges on them: The refereed literature will already be free for everyone as a result of steps *i-iv*, irrespective of the outcome of predictions *v-viii*.

#### *i. Universities install and register OAI-compliant Eprint Archives (<http://www.eprints.org>).*

The Eprints software is free and will be open-sourced. It in turn uses only free software; it is quick and easy to install and maintain; it is OAI-compliant and will be kept compliant with every OAI upgrade: <http://www.openarchives.org/>. Eprints Archives are all interoperable with one another and can hence be harvested and searched (e.g., <http://arc.cs.odu.edu/>) as if they were all in one global "virtual" archive of the entire research literature, both pre- and post-refereeing.

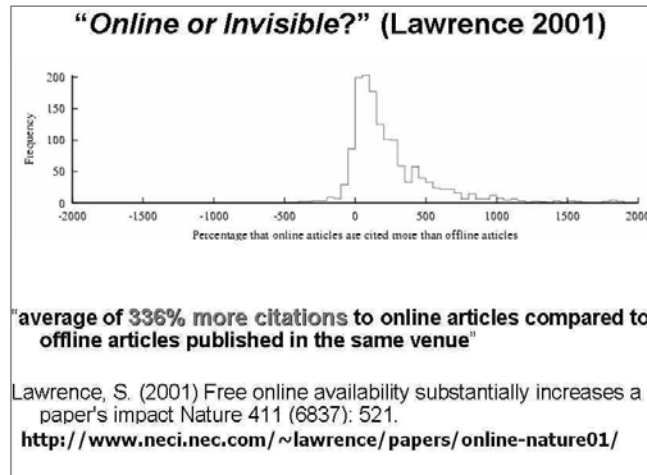
#### *ii. Authors self-archive their pre-refereeing preprints and post-refereeing postprints in their own university's Eprint Archives.*

This is the most important step; it is insufficient to create the Eprint Archives. All researchers must self-archive their papers therein if the literature is to be freed of its access- and impact-barriers. Self-archiving is quick and easy; it need only be done once per paper, and the result is permanent, and permanently and automatically uploadable to upgrades of the Eprint Archives and the OAI-protocol.

#### *iii. Universities subsidize a first start-up wave of self-archiving by proxy where needed.*

Self-archiving is quick and easy, but there is no need for it to be held back if any researcher feels too busy, tired, old or otherwise unable to do it for himself: Library staff or students can be paid to "self-archive" the first wave of papers by proxy on their behalf. The cost will be negligibly low per paper, and the benefits will be huge; moreover, there will be no need for a second wave of help once the palpable benefits (access and impact) of freeing the literature begin to be felt by the research community. Self-archiving will become second-

nature to all researchers as the objective digitometric indicators of its effects on citations and usage become available online (Harnad 2001e; Lawrence 2001a, 2001b) (e.g., cite-base or ResearchIndex).



*iv. The give-away corpus is freed from all access /  
 impact barriers on-line.*

Once a critical mass of researchers has self-archived, the refereed research literature is at last free of all access- and impact-barriers, as it was always destined to be.

## 4.2 Hypothetical sequel

Steps *i-iv* are sufficient to free the refereed research literature. We can also guess at what may happen after that, but these are really just guesses. Nor does anything depend on their being correct. For even if there is no change whatsoever – even if Universities continue to spend exactly the same amounts on their access-toll budgets as they do now – the refereed literature will have been freed of all access/impact barriers forever.

However, it is likely that there will be some changes as a consequence of the freeing of the literature by author/institution self-archiving. This is what those changes might be:

*v. Users will prefer the free version?*

It is likely that once a free, online version of the refereed research literature is available, not only those researchers who could not access it at all before, because of toll-barriers at their institution, but virtually all researchers will prefer to use the free online versions.

Note that it is quite possible that there will always continue to be a market for the toll-based options (on-paper version, publisher's on-line PDF, deluxe enhancements) even though most users use the free versions. Nothing hangs on this.

*vi. Publisher toll revenues shrink, library toll savings grow?*

But if researchers do prefer to use the free online literature, it is possible that libraries may begin to cancel journals, and as their windfall toll savings grow, journal publisher toll revenues will shrink. The extent of the cancellation will depend on the extent to which there remains a market for the toll-based add-ons, and for how long.

If the toll-access market stays large enough, nothing else need change.

*vii. Publishers downsize to providers of peer-review service and optional add-ons products?*

It will depend entirely on the size of the remaining market for the toll-based options whether and to what extent journal publishers will have to down-size to providing only the essentials: The only essential, indispensable service is peer review.

*viii. Peer-review service costs funded by author-institution out of reader-institution toll savings?*

If publishers can continue to cover costs and make a decent profit from the toll-based *optional add-ons* market, without needing to down-size to peer-review provision alone, nothing much changes.

But if publishers do need to abandon providing the toll-based products and to scale down instead to providing only the peer-review service, then universities, having saved 100% of their annual access-toll budgets, will have plenty of annual windfall savings from which to pay for their own researchers' continuing (and essential) annual journal-submission peer-review costs (10-30%); the rest of their savings (70-90%) they can spend as they like (e.g., on books – plus a bit for Eprint Archive maintenance).

## 5. PostGutenberg Copyright Concerns

There is a great deal of concern about copyright in the digital age, and some of it may not be easily resolvable (e.g., what to do about the pirating of software and music). But none of that need detain us here, because digital piracy is only a problem for *non-give-away* work, whereas we are concerned here only with *give-away* work. (Again, failing to make the give-away/non-give-away distinction leads only to confusion, and the *misapplication* of the much bigger and more representative *non-give-away* model to the anomalous give-away corpus, which it does not fit.)

The following digital copyright concerns are *relevant to the non-give-away literature only*:

### 5.1 Protecting intellectual property (royalties)

This is as much of a concern to authors of books as to authors of screenplays, music, and computer programs. It is also a concern to performers who have made digital audio or video disks of their work. They do not wish to see that work stolen; they want their fair share of the gate-receipts in return for their talent and efforts in producing the work.

But the producers of refereed research reports do not wish to have protection from "theft" of this kind; on the contrary, they wish to *encourage* it. They have no royalties to gain from preventing it; they have only research impact to *lose* from access-blockage of any kind.

### 5.2 Allowing fair use (user issue)

"Fair use" is another worthy concern. It has to do with certain sanctioned uses of *non-give-away* material, such as all or parts of books, magazine articles, etc., often for teaching purposes; the producers of these works do *not* wish to lose their potential royalty/fee-income from these works.

The producers of refereed research reports, in contrast, wish to give their work away; hence fair-use issues are moot for this special give-away literature.

### 5.3 Preventing theft of text (piracy)

The producers of refereed research reports do not wish to prevent the theft of their texts; they wish to facilitate it as much as possible. (In the on-paper era they used to purchase and mail reprints to requesters at their own expense!)

The following digital copyright concern is relevant to *all* literature, both give-away and non-give-away:

#### 5.4 Preventing theft of authorship (plagiarism)

No author wants any other author to claim to have been the author of his work. This concern is shared by all authors, give-away and non-give-away. But it has *nothing whatsoever* to do with concerns about theft-of-text, and should not be conflated with such concerns in any way: Give-away work need not be held hostage to non-give-away concerns about theft-of-text under the pretext of "protecting" it from theft-of-authorship. (Unfortunately, many journal publishers try to write and use their copyright transfer agreements for precisely this purpose, and authors need to become aware of it.)

The following digital copyright concern is relevant to the *give-away* literature only:

#### 5.5 Guaranteeing author give-away rights

Apart from the protection from plagiarism and the assurance of priority that all authors seek, the only other "protection" the give-away author of refereed research reports seeks is protection of his give-away rights!

(The intuitive model for this is advertisements: what advertiser wants to lose his right to give away his ads for free, diminishing their potential impact by charging for access to them!)

Well, there is no need for the authors of refereed research to worry about exercising their give-away rights, for they can do it, legally, even under the most restrictive copyright agreement, by using the following strategy.

### 6. How to Get Around Restrictive Copyright Legally ("Preprint+Corrigenda Strategy")

#### 6.1 Self-archive the pre-refereeing preprint

Self-archiving the preprint is the critical first step. Before it has even been submitted to a journal, your intellectual property is your own, and not bound by any future copyright transfer agreement. So archive the preprints (as physicists have done for 10 years now, with over 150,000 papers, and cognitive scientists have done for 3 years now, with over 1000 papers). This is a good way to establish priority, elicit informal feedback, and keep a public record of the embryology of knowledge.

[Note that some journals have, apart from copyright policies, which are a legal matter, embargo policies," which are merely policy matters (nonlegal). Invoking the "Ingelfinger (Embargo) Rule," some journals state that they will not *referee* (let alone publish) papers that have previously been "made public" in any way, whether through conferences, press releases, or on-line self-archiving. The Ingelfinger Rule, apart from being directly at odds with the interests of research and researchers and having no intrinsic justification whatsoever – other than as a way of protecting journals' current revenue streams – is not a legal matter, and unenforceable. So researchers are best advised to ignore it completely (Harnad 2000a, 2000b), exactly as the authors of the 150,000 papers in the Physics Archive have been doing for ten years now. The Ingelfinger Rule is under review by journals in any case; Nature has already dropped it, and there are indications that Science may soon follow suit too.]

## 6.2 Submit the preprint for refereeing (revise etc.)

Nothing changes in author publication practises; nothing needs to be given up. Submit your preprint to the refereed journal of your choice, and revise it as usual in accordance with the directive of the Editor and the advice of the referees.

## 6.3 At acceptance, try to fix the copyright transfer agreement to allow self-archiving

Copyright transfer agreements take many forms. Whatever the wording is, if it does not explicitly permit online self-archiving, modify it so that it does. Here is a sample way to word it (<http://cogprints.soton.ac.uk/copyright.html>):

I hereby transfer to [publisher or journal] all rights to sell or lease the text (on-paper and on-line) of my paper [paper-title]. I retain only the right to distribute it for free for scholarly/scientific purposes, in particular, the right to self-archive it publicly online on the Web.

Some publishers (about 10-30%) already explicitly allow self-archiving of the refereed postprint (e.g., the American Physical Society: <http://forms.aps.org/author/copytrnsfr.pdf>). Most other publishers (perhaps 70%) also accept this clause, but only if you explicitly propose it yourself (they will not formulate it on their own initiative).

#### 6.4 If 6.3 is successful, self-archive the refereed postprint

Hence, for about 80% of journals, once you have done the above, you can go ahead and self-archive your paper.

Some journals (about 20%), however, will respond that they decline to publish your paper unless you sign their copyright transfer agreement verbatim. In such cases, sign their agreement and proceed to the next step:

#### 6.5 If 6.3 is unsuccessful, archive the "corrigenda"

Your pre-refereeing preprint has already been self-archived since prior to submission, and is not covered by the copyright agreement, which pertains to the revised final ("value-added") draft. Hence all you need to do is to self-archive a further file, linked to the archived preprint, which simply lists the corrections that the reader may wish to make in order to conform the preprint to the refereed, accepted version.

Everyone chuckles at this point, but the reason it is so easy is that this is the author give-away literature. No *non-give-away* author would ever dream of doing such a thing (archiving the prepublication draft for free, along with the corrigenda). And copyright agreements (and copyright law) are designed and conceived to meet the much more representative interests of non-give-away authors and their much larger body of royalty/fee-based work. Hence this simple and legal expedient for the special, tiny, anomalous, give-away literature has no constituency anywhere else.

Yet this simple, risible strategy is also feasible, and legal (Oppenheim 2001) – and sufficient to free the entire current refereed corpus of all access/impact barriers immediately!

### 7. What *You* Can Do *Now* to Free the Refereed Literature Online

#### 7.1 *Researchers*: Self-archive all present, future (& past) papers

The freeing of their present and future refereed research from all access- and impact-barriers forever is now entirely in the hands of researchers. Posterity is looking over our shoulders, and will not judge us flatteringly if we continue to delay the optimal and inevitable needlessly, now that it is clearly within our reach. Physicists have already shown the way, but at their current self-archiving rate, even they will take another decade to free the entire Physics literature (<http://www.cogsci.soton.ac.uk/~harnad/Tp/Tim/sld002.htm>) – with

the Cognitive Sciences (<http://cogprints.soton.ac.uk>) 39 times slower still, and most of the remaining disciplines not even started: <http://www.cogsci.soton.ac.uk/~harnad/Tp/Tim/sld004.htm>.

This is why it is hoped that (with the help of the eprints.org institutional archive-creating software) distributed, institution-based self-archiving, as a powerful and natural complement to central, discipline-based self-archiving, will now broaden and accelerate the self-archiving initiative, putting us all over the top at last, with the entire distributed corpus integrated by the glue of *interoperability* (<http://www.openarchives.org>).

As to the past (retrospective) literature: The preprint+corrigenda strategy will not work there, but as the retrospective journal literature brings virtually no revenue, most publishers will agree to author self-archiving after a sufficient period (six months to two years) has elapsed. Moreover, for the really old literature, it is not clear that on-line self-archiving was covered by the old copyright agreements at all.

And if all else fails for the retrospective literature, a variant of the preprint+corrigenda strategy will still work: Simply do a revised 2nd edition! Update the references, rearrange the text (and add more text and data if you wish). For the record, the enhanced draft can be accompanied by a "de-corrigenda" file, stating which of the enhancements were *not* in the published version.

(And of course the starting point for the revised, enhanced 2nd edition, if you no longer have the digital text in your word processor, can be scanned and OCR'd from the journal; by thus distributing it, authors can do for their own work for-free what JSTOR <http://www.jstor.org/> is only able to do for the work of others for-fee.)

## 7.2 Universities: Install Eprint Archives, mandate them; help in author start-up

Universities should create institutional Eprint Archives (e.g., CalTech) for all their researchers. They should also mandate that they be filled. <http://www.eprints.org/signup/sign.php> It is already becoming normal practise for faculty to keep and update their institutional CVs online on the Web; it should be made standard practise that all CV entries for refereed journal articles are linked to their archived full-text version in the university's Eprint Archive.

For researchers who profess to be too busy, tired, old, or inexpert to self-archive their papers for themselves, a modest start-up budget to pay library experts or students to do it for them would be a small amount of money very well-invested. It will only be needed to get the first wave over the top; from then on, the momentum from the enhanced access and impact will maintain itself, and self-archiving will become as standard a practise as email.



But what needs energetic initial promotion and support is the first wave. If (i) the enhanced access of their own researchers to the research of others and (ii) the enhanced visibility (Lawrence 2001a, 2001b) and the resulting enhanced impact of their own research on the research of others are not incentive enough for universities to promote and support the self-archiving initiative energetically, they should also consider that it will be an investment in (iii) a potential solution to their serials crisis and the possible recovery of 70-90% of their annual serials (toll-access) budget.

(Note that the success of the self-archiving initiative is predicated on the same Golden Rule on which both refereeing and research themselves are predicated: If we all do our own part for one another, we all benefit from it. Give in order to receive ...)

### 7.3 *Libraries*: Maintain the University Eprint Archives; help in author start-up

Libraries are the most natural allies of researchers in the self-archiving initiative to free the refereed journal literature. Not only are they groaning under the yoke of the growing serials budget crisis, but librarians are also eager to establish a new digital niche for themselves, once the journal corpus is on-line: Maintaining the Eprint Archives, and facilitating the all-important start-up wave of self-archiving (by being ready to do "proxy" self-archiving on behalf of authors who feel they cannot do it for themselves), will be a critical role for libraries to play.

Libraries can also facilitate a stable transition through their collective, consortial power (SPARC: <http://www.arl.org/sparc>), providing leveraged support for publishers who are prepared to commit themselves to a scheduled for downsizing to the essentials only (the peer review service, to the author/-institution). And individually they can also be preparing in advance for the restructuring that will come if their windfall toll savings grow; about 10-30% of their annual savings will need to be redirected to cover their university's own authors' peer-review charges per paper. The remaining 70-90% is theirs to use in any way they see fit!

### 7.4 *Students*: *Stay the course! Surf!* The future is optimal, inevitable and yours!

Students are well-advised to keep doing what they do naturally: Favor material that is freely accessible on the Web. This will not net them very much of the *non-give-away* literature, but it will put consumer pressure on the give-away research literature, especially as these students come of age, and become researchers in their turn.

*7.5 Publishers: Concede graciously on self-archiving and be prepared to separate essential peer-review service costs (to the author-institution) from optional add-on product costs (to the reader-institution)*

Publishers should concede graciously on self-archiving as the American Physical Society (APS) has done and not try to use copyright or embargo policy to prevent or retard it. Such measures are in direct conflict with the interests of research and researchers, they are destined to fail, they can already be legally circumvented, and they only make publishers look bad.

A much better policy is to concede on the optimal and inevitable for research, and plan on the possibility of separating the provision of the essential peer review *service* to the author-institution (peer review implementation charges, per paper) from the provision of all other add-on *products* (e.g., on-paper version, on-line version, other added-values), which should be sold as options, rather than used to try to keep holding the essentials (the refereed final draft) hostage to access-tolls.

There will still be a permanent niche for journal publishers. What remains to be seen is whether that will entail downsizing to peer-review service-provision alone, or whether there will also continue to be a market for toll-based add-ons even after the refereed drafts are available free through the Eprint Archives.

*7.6 Government/society: Mandate public archiving of public research worldwide*

Government and society should support the self-archiving initiative, reminding themselves that most of this giveaway research has been supported by public funds, with the support explicitly conditional on making the research findings public (<http://www.sciencemag.org/cgi/content/full/281/5382/1459>). In the PostGutenberg Galaxy there is no longer any need for that public accessibility to be blocked by toll-barriers.

The beneficiaries will not just be research and researchers, but society itself, inasmuch as research is supported because of its potential benefits to society. Researchers in developing countries and at the less affluent universities and research institutions of developed countries will benefit even more from barrier-free access to the research literature than will the better-off institutions, but it is instructive to remind ourselves that even the most affluent institutional libraries cannot afford most of the refereed journals! None have access to more than a small subset of the entire annual corpus (<http://fisher.lib.virginia.edu/ar1/index.html>). So free access to it all will benefit all institutions (Odlyzko 1999a, 1999b).

And on the other side of barrier-free access *to* the work of others, all researchers, even the most affluent, will benefit from the barrier-free impact of

their own work *on* the work of others. Moreover, a freed, interoperable, digital research literature will not only radically enhance access, navigation (e.g., citation-linking) and impact, hence research productivity and quality, but it will also spawn new ways of monitoring and measuring that impact, productivity and quality (e.g., download impact, links, immediacy, comments, and the higher-order dynamics of a citation-linked corpus that can be analyzed from preprint to post-postprint, to yield an "embryology of knowledge" (Harnad & Carr 2000).

## 8. Prima-Facie FAQs for Overcoming Zeno's Paralysis

"I worry about self-archiving because ...":

Researchers, librarians, publishers and university administrators have so far been held back from self-archiving by certain prima facie worries, all of which are easily shown to be groundless.

These worries are rather like "Zeno's Paradox":

"I cannot walk across this room, because before I can walk across it, I must first walk half-way across it, and that takes time; but before I can walk half-way across it, I must walk half-half-way across it, and that too takes time; and so on; so I how can I ever even get started?"

This condition might better be called "Zeno's Paralysis."

Each of the following worries can easily be shown to be groundless (and *has* been shown to be groundless, by myself and many others, many times). Yet the very same prima facie worries keep resurging elsewhere, like mushrooms, no matter how decisively they are uprooted in each instance. It will be a matter for future historians to explain the puzzle of why we were needlessly held back for so long from the optimal and inevitable even when it was well within reach, by these gratuitous worries (despite the "Los Alamos Lemma," which is that whatever alleged obstacle was not sufficient to deter physicists from self-archiving 130,000 papers to date should not be holding back the rest of us either!).

Here are rebuttals to the most common of these prima facie worries; in future they can be used as FAQs to reply by number: They are brief and to the point, because there are no long, complex, hidden issues in any of these cases. Hence it is best to get to the point in the simplest, most direct way possible. There is also a good deal of overlap and redundancy between them:

### 8.1 Preservation

"I worry about self-archiving because archived eprints may not continue to exist or to be accessible *in perpetuo* on-line, the way they were on-paper."

To put this worry into perspective, we must remember that print-on-paper is not permanent either. The only relevant parameter is the *probability* of future access. The on-paper probability, such as it is, is achieved by generating (a) multiple copies, (b) geographically distributed, (c) in a (relatively) robust medium, (d) visible to the human eye.

All four of these properties can be (and have been) achieved on-line too, and the resulting preservation probability can be made as good as, or even better than, the current probability on-paper.

That should be the end of the story: For once this concern is no longer grounded in actual, objective probabilities, but only in prior habits and attendant intuitions, then we are talking about biases and superstitions and not about actual risks.

There are a few side issues: People worry about global power-failures, or global dictatorships. They should remind themselves that these are matters of probability too, and have their equivalents in paper.

People also, by analogy with current unreadable documents in obsolete word-processors or peripherals, worry about whether the digital code, even if preserved, will always be accessible and visible to the eye.

The answer is again probability: The reason print-on-paper has been faithfully preserved across generations (when it has been) is that the literate world's collective interests were vested in ensuring that it should do so. This same continuity of collective interests will exist for the digital corpus too, for the same reasons, except that digital code will be much *easier* to keep uploading to every successive new technology than print on-paper to every successive building or regime ever was.

(And there is always the option for those who are still not confident enough in the technology, despite the odds, of printing out hard copies as back-up: Indeed, that is a good way to put the magnitude of one's Zeno's worries to the test: Who will still feel the need to make hard copies, and of how much of the corpus, once it's all on-line and accessible to everyone, everywhere, at all times?)

In short, preservation measures as a practical pursuit by digital librarians is an eminently worthy one; but as a basis for any hesitation or delay whatsoever about proceeding with self-archiving right now, it is completely irrational (particularly as, for the time being, self-archiving is merely a *supplement* to, not a *substitute* for, the existing Gutenberg modes of preservation).

## 8.2 Authentication

"I worry about self-archiving because you can never be sure whether you are reading the definitive version of an eprint on-line, the way you can be sure on-paper."

Again, the rational way to put this into context and proportion is to remind ourselves that the authenticity of an on-paper version is just a matter of probability too, and that the very same factors that maximize that probability on-paper can maximize it on-line too. Indeed, if we wish, we can make both the probability and the verifiability of authenticity on-line much higher than it currently is on-paper through techniques such as public hash/time-stamping and encryption.

Nor should the authentication issue be confused with the issue of peer-review (8.7) or journal certification (8.5) (separate questions), nor with the question of "version control" (there will be self-archived preprints, revised drafts, final accepted, published drafts [postprints], updated, corrected post-postprints, peer comments, author replies, revised second editions. In all of this, the refereed, accepted final draft is one crucial "milestone," but not the only one, in the embryology of knowledge [and not even always the best one]).

And last, some of the "authentication" worries arise from conflating self-archiving and self-publication. To say it in longhand: The main objective of the self-archiving initiative is the freeing of the *refereed* drafts from access/impact barriers. The refereed draft has already been "authenticated" by the journal that peer-reviewed it. Do not confuse that authentication with some worry you may have about whether this self-archived draft is indeed what the author purports it to be. The only thing the author is "self-certifying" in this case is that this is indeed the *journal-certified* final draft. There is of course always a possibility that it is *not* the journal-certified final draft; but that was also true when the author sent you an on-paper reprint. The probabilities can, as usual, be tightened to make them as high as we feel comfortable with in either case. And in the case of preservation, self-archiving is at this stage merely a supplement, not a substitute for existing forms of authentication.

So, again, there are no rational authentication concerns whatsoever to deter us from self-archiving immediately.

### 8.3 Corruption

"I worry about self-archiving because eprints can be altered or otherwise corrupted on-line in ways they could not be corrupted on-paper."

If the "authentication" worry (8.2) is the worry about "self-corruption" by the author who has self-archived his own paper, this second "corruption" worry is about "allo-corruption" by parties other than the author.

Again, the answer is that simple and effective means are available to ensure that an on-line draft is uncorrupted with as high a probability as we feel we need. So this too is a non-problem. (Nor should it, again, be conflated with self-publication issues, which are irrelevant to the self-archiving of refereed, journal-published papers.) Whatever level of incorruptibility we feel we need, we can have it for self-archived papers too.

Consequently, corruptibility worries provide no rational basis whatsoever for deterring us from self-archiving immediately.

#### 8.4 Navigation (info-glut)

"I worry about self-archiving because there is already too much to read, and it is already too hard to navigate it on paper; adding eprints will just make this situation even worse."

This worry deserves even less space than the others. It is incontestable that the information glut (<http://www.sims.berkeley.edu/how-much-info/summary.html>) is far more navigable and manageable on-line than on-paper (Harnad & Carr 2000).

The primary objective of self-archiving is to free the refereed journal literature from access-tolls on-line. That literature is *already* being published on-paper. (If you think it should not be, it is with the journals and their referees that you need to take issue, not with self-archiving or the on-line medium!) When it is all accessible free on-line, there is no need for anyone to feel any more (or less) obliged to read the refereed literature than they did on-paper. Keeping it off-line is certainly no cure for the information glut (if there is one); it merely makes the existing access-tolls the arbitrary arbiters of whether or not one reads something, rather than the reader's own rational judgement. (And unrefereed preprints can of course always be ignored altogether, if the reader wishes, on-line just as on-paper.)

In short, no rational deterrent at all to immediate self-archiving from concerns about navigation or information glut.

#### 8.5 Certification

"I worry about self-archiving because papers are not certified on-line, the way they are in a journal on-paper."

This worry is again based on conflating publication and archiving: The journal publisher (and referees) certify; the archive merely provides access. The author, in self-archiving, "self-certifies" his refereed, published draft as indeed being the one that the journal refereed and published. And this being the case is, as usual, a matter of probability, whether on-line or on-paper. And that probability can be made as high as we feel we need.

Again, no rational deterrent to immediate self-archiving in the certification worry.

#### 8.6 Evaluation

"I worry about self-archiving because there is no evaluative process on-line as there is on-paper."

Again, a conflation of publishing and archiving: Journal editors and their referees evaluate drafts and revisions, and if/when they are satisfied that their journal's quality standards have been met, they certify the final draft as having met them (peer review). The author self-archives the peer-reviewed postprints (and unrefereed preprints, and perhaps revised post-postprints), tagging them correspondingly. We can decide how high a probability we need that the peer-reviewed draft is indeed the peer-reviewed draft, but that is not the problem of evaluation, but just the question of authentication (8.2) again.

So there is no rational deterrent to immediate self-archiving anywhere in the evaluation worry.

### 8.7 Peer review

"I worry about self-archiving because on-line eprints are not refereed, as they are on-paper: What will become of peer review?"

Again, a conflation of publishing and archiving: The author self-archives both pre-refereeing preprints and refereed postprints (etc.). The peer review continues to be performed by the referees, as it always was. Peer-review is medium-independent.

No rational deterrent to immediate self-archiving in the peer-review worry.

### 8.8 Paying the piper

"I worry about self-archiving because someone surely has to pay for all this: you can't get something for nothing!"

There are many fallacies embedded in this worry, among them misunderstandings about the nature of global networked communication. Internet connectivity, at very low cost, is now part of the infrastructure of most of the world's universities and research institutions. If you are not equally worried about who pays for your emails, websites, and web-browsing, you should not be worrying about your self-archiving either. In any case, paying access-tolls is not paying the pertinent piper here anyway!

The refereed research literature is minuscule compared to the rest of the traffic on the Web (<http://www.sims.berkeley.edu/how-much-info/summary.html>). It is the flea on the tail of the dog. Worry about the storage and band-width for the growing daily creation and use of audio, video, and multimedia (most of it non-research use!) by researchers at universities and research institutions before even beginning to fret about the refereed flea. As usual, there is also some of the archiving/publishing conflation here, thinking that we must find some sort of counterpart for the printing/distribution costs, somewhere. But there isn't any. The price per-paper of permanent online archiving is virtually zero, yet everyone, everywhere, has access to it all, forever. This is a Gutenberg expense

that has simply vanished in the PostGutenberg Galaxy, leaving only the Cheshire Cat's Grin.

There is indeed one essential publishing cost that still needs to be paid, but it has nothing to do with Internet use: It is the cost of implementing peer review. That cost, however, as discussed in the peer review section (3.3), is only 10-30% of the access-toll costs currently being paid, and hence could easily be paid out of the annual savings.

The last of the "who-pays-the-piper" worries is, I think, a variant of the capitalism (8.14) worry. The best way to dispel it is to note that refereed publishing in the PostGutenberg Galaxy, once the literature has been freed through self-archiving, is likely (apart from whatever optional add-on products and services there may still be a market for) to downsize into a *service* (peer review), provided to the author-institution, instead of the toll-based *product* (the text) that was provided to the reader-institution in the Gutenberg era.

Nothing hinges on this, however, for as long as the world wants to keep paying for the toll-based product, even after the refereed literature has been self-archived, the piper will be fully paid, yet the literature will be free of all its access/impact barriers.

No rational deterrent to immediate self-archiving in the who-pays-the-paper worries.

## 8.9 Downsizing

"I worry about self-archiving because it may force journal publishers to shrink to a non-sustainable size, and then where would we be?"

No one can predict with certainty the evolutionary path that scientific/scholarly journal publishing will take once the refereed corpus has been freed online by self-archiving. The toll-based market for the on-paper version, for the publisher's on-line version or for other options may continue indefinitely, or it might shrink but re-stabilize at a lower level, or it might disappear altogether – and this could happen relatively slowly or relatively quickly.

It is not clear in advance which of the current established journal publishers will want to continue doing what, under what conditions. The bottom line is that the only remaining *essential* service will be peer review. If and when that is the only service for which there remains a market, either current journal publishers will be able and willing to downsize to that niche, or they will terminate journal operations, in which case their titles (that is, each journal's editor, editorial board, referees, and authorship) will simply migrate to new on-line only journal publishers who are ready to adapt to the new niche [e.g., the Institute of Physics's New Journal of Physics (<http://www.njp.org/>) and Bio-Med Central (<http://www.biomedcentral.com/>)].

No rational deterrent to immediate self-archiving in worries about publisher downsizing.



## 8.10 Copyright

"I worry about self-archiving because it is illegal, it violates copyright agreements, and can jeopardize career and livelihood."

Please see the sections on copyright and on legal ways to self-archive despite restrictive copyright transfer agreements.

In brief, many journals will agree to author self-archiving if the author asks, and for those that don't, self-archiving the preprint before submission and a "corrigenda" file after acceptance is sufficient, and completely legal. What career and livelihood depend on is peer review and impact, and all self-archiving authors continue to have both; neither needs to be sacrificed for the other.

No rational deterrent to immediate self-archiving in copyright worries.

## 8.11 Plagiarism

"I worry about self-archiving because it is so much easier to steal someone else's text on-line, and publish it as one's own, than it is to do so on-paper."

This is again a matter of probability: Yes, "it is much easier to steal someone else's text on-line, and publish it as one's own, than it is to do so on-paper," but it is also much easier to detect such thefts on-line; and it is possible to do both (steal and detect) on-paper too.

Depending on how important we find it to do so, we can make escape from detection so improbable on-line that it becomes harder to plagiarize on-line than on-paper. It is not clear, however, whether it is even all that important to do so. Worries about plagiarism are usual based on the archiving/publishing conflation: Once one's findings have been refereed and published, it is hard for anyone else to derive any benefit from them at the expense of the author (the peer-reviewed version settles all subsequent authorship disputes).

Pre-refereeing preprints are another story; they are dealt with partly in the prior discussion of authentication (8.2), and partly under priority (8.12), below.

For refereed postprints, however, refraining from self-archiving them because of worries about plagiarism would be no more rational than refraining from publishing them on-paper in the first place, for the very same reason.

## 8.12 Priority

"I worry about self-archiving because one cannot establish priority on-line as one can on-paper."

Establishing priority is again a matter of probability, but it can readily be made much more definitive and reliable on-line than on-paper if we wish. See authentication (8.2).

No rational deterrent to immediate self-archiving in priority worries.

### 8.13 Censorship

"I worry about self-archiving because censors could decide what can and cannot appear on-line."

This worry too is probably based in part on the usual archiving/publishing conflation (casting the Web and the archive in the role of a publisher who refuses to publish your work).

It is true that one's on-line literary goods are at the mercy of the archives and archivists. But one's analog on-paper literary goods were likewise at the mercy of the libraries. They could have chosen to "censor" our work too.

Again, it is just a matter of deciding how tight we wish to make the probabilities in this medium. Mirroring, caching/harvesting and distributed coding already go some way toward taking it out of any potentially sinister local hands.

No rational deterrent to immediate self-archiving in worries about censorship.

### 8.14 Capitalism

"I worry about self-archiving because access-tolls are hallmarks of capitalism, market economics, supply and demand, free enterprise. Give-aways smack either of socialism, or market interference, or non-sustainability."

This too is merely a superstition. There are plenty of perfectly capitalistic precedents for give-aways, *advertising* being the most prominent one. If the thought of advertisers curtailing the potential impact of their ads by charging potential customers for access to them makes no sense, then it makes just as little sense to curtail the potential impact of research by charging potential users for access to them.

Nor is there any market interference in self-archiving one's own refereed research: If institutions and individuals want to pay for toll-access to the on-paper version, or the publisher's PDF, or further options, they can still do so; but there is no longer any need or justification for continuing to hold the essentials (the peer-reviewed draft) hostage to those toll-based options in the PostGutenberg era, any more than there was any need or justification for continuing to hold the essentials of long-distance communication hostage to postal transport costs in the era of telephony. (Rather than capitalism being under assault from self-archiving, trying to prevent researchers from benefiting from this new, more efficient and economical way of disseminating and maximizing the impact of their refereed research smacks of protectionism.)

Two variants on the capitalism-worry arise from scepticism about the eventual transition from providing an toll-based product to the reader-institution to providing a peer-review service to the author-institution. Note that, strictly speaking, it is not even necessary to answer these worries, as this eventual

transition is hypothetical, whereas freeing the refereed literature now through self-archiving is not; but here are replies anyway:

Question 1: "Won't paying directly for the peer-review service lead to inflated peer-review costs by the most prestigious journals?"

Question 2: "Won't peer-review revenues lower standards, so that lower-quality work is accepted in order to get more peer-review revenue?"

The answer to both is similar: Referees referee for free, and journal quality and prestige (and impact) depend on *rejection* rates. Trying to inflate revenue by lowering acceptance thresholds simply lowers quality, thereby favoring the competition, with higher standards. It is a built in counter-weight. Likewise for raising peer-review rates: As referees referee for free, there is no reason one journal should charge more than another, and if they do, they risk driving not only the authors but the unpaid referees to the competition. Because the competitive commodity in this anomalous give-away domain is quality, and nothing else.

A proposal has occasionally been voiced to keep preserve toll-barriers by buying authors off from self-archiving, by offering to share the revenue with them (royalty payments). But the trade-off between imprint-income and impact-income is so disproportionate for this anomalous domain that there is not faintly enough money available to make others prefer sacrificing their potential impact in exchange.

No rational deterrent to immediate self-archiving in worries about capitalism.

### 8.15 Readability

"I worry about self-archiving because it is inconvenient to read texts on screen, and hard on the eyes. It is also not suitable for bed, beach or bathroom reading."

At the moment it is undeniable that for extended, discursive reading, on-paper is still preferable to on-line. This will no doubt change, but even now it is no reason whatsoever for not self-archiving. First, a large proportion of the scientific and scholarly use of the refereed research literature consists of browsing and searching, not linear reading, and for this, on-line navigation is already incomparably superior. Second, there is still that vast potential readership to consider, whose access to your research in *any* form is currently blocked by unaffordable access-tolls (Odlyzko 1999a, 1999b; <http://www.arl.org/stats/index.html>); for that entire disenfranchised population, it's either online or not at all. And last, even for linear reading, the archived version can always be printed off.

No rational deterrent to immediate self-archiving in worries about readability.

## 8.16 Graphics

"I worry about self-archiving because on-line graphics have coarser resolution than on-paper and require too much storage capacity and transmission time."

Graphics too will no doubt improve. With a few exceptions, such as fine arts and histology, digital graphics are already good enough. Users can always decide whether or not they feel they need to access the deluxe hard copy; no need to make a pre-emptive decision on their behalf, as the on-line version is in any case a supplement, not a substitute, for the time being. And graphics are quite a natural test-bed to see whether there is still any market left for any for-fee add-ons.

No rational deterrent to immediate self-archiving in worries about graphics.

## 8.17 Publishers' future

"I worry about self-archiving because of what it might do to journal publishers' future."

See the replies about paying the piper (8.8), downsizing (8.9), and capitalism (8.14). Those journal publishers who are willing and able to scale down to their new PostGutenberg niche can do so. New online-only journal publishers are ready to take over the titles in the cases where they are not. The remaining peer-review service costs per submitted paper can be paid for by the author-institution out of 10-30% of its annual 100% toll savings. And refereed journal publication is only a small portion of publication, most of the rest of which, being *non-give-away*, will proceed on-line much the way it does on-paper.

No rational deterrent to immediate self-archiving in worries about publishers' future.

## 8.18 Libraries'/Librarians' future

"I worry about self-archiving because of what it might do to libraries' and librarians' future."

The serials literature is all going on-line anyway, irrespective of the speed or success of the self-archiving initiative. If this requires restructuring of some librarian skills and functions, this will take place in any case. Some have thought that managing digital serials collections will fill the gap, but it is not clear how much management those will need, apart from paying the annual toll-bills! Author/institution Eprint Archives, on the other hand, will call for more digital librarian skills, in everything from helping researchers to do the self-archiving, to maintaining the institution's Eprint Archive and seeing to its continued interoperability with the rest of the world's Eprint Archives, its upgrading, and its preservation.

Moreover, in implementing and maintaining the institutional Eprint Archives, Libraries will be investing in the solution of their serials crisis. Of the 100% annual toll budget that this can potentially save, after 10-30% of it has been redirected to cover author-institution peer-review costs, the remaining 70-90% can be used to fund other librarians' activities, including the purchase of non-give-away materials such as books (whether on-paper or on-line).

No rational deterrent to immediate self-archiving in worries about libraries'/librarians future.

### 8.19 Learned Societies' future

"I worry about self-archiving because of what it might do to Learned Societies' future."

Learned Societies are potential allies in and beneficiaries of the self-archiving initiative. First, they are *us*. Whatever is good for research, and for research impact, is therefore good for Learned Societies.

But many of them are also journal publishers, and hence may be facing downsizing pains. Unlike commercial publishers, however, their first and last allegiance will of course be to research and researchers, that is, us. We will hear rationalizations about needing the toll revenues to fund "good works" such as meetings, scholarships and lobbying. But it will quickly become evident that, on the one hand, some of these good works are not essentials either, and certainly nothing that we would want to sacrifice research impact for; and the subset of them that really is essential (such as meetings) will prove to be able to fund itself other ways too, rather than needing to be subsidized at the expense of research impact.

Learned Societies (and perhaps also University Presses) are also natural candidates for taking over the serials titles of commercial journal publishers who prefer to discontinue journal operations rather than scale down to just becoming peer-review service providers.

No rational deterrent to immediate self-archiving in worries about Learned Societies' future.

### 8.20 University conspiracy

"I worry about self-archiving because I worry that universities may have other plans for their researchers' writings, such as Eprint Archive access-tolls."

This worry seems to be based on some (one hopes) over-suspicious views about university administrators and their motives.

We should not forget that the give-away refereed literature is esoteric, with virtually no "market" per paper. So whereas there might be a basis for suspicion about what our hard-pressed universities might like to do if they could get their hands on our exoteric, *non*-give-away work (royalty-bearing books and

textbooks), there's not much they could do to squeeze revenue out of our no-market, give-away refereed research reports even if they wanted to. On the contrary, our universities, like ourselves, benefit far more from the potential impact-income of such work – maximized by removing all access-barriers – than from any potential imprint-income that could be squeezed out of it by co-opting the "P" from the publishers' S/L/P tolls and using it to charge institutional archive access-tolls.

Moreover, our universities' potential toll savings, and relief from their serials crises, are completely dependent on freeing access to our research. Any sign of university-levied archive-access tolls would simply serve to keep the current access-tolls in place.

No rational deterrent to immediate self-archiving in worries about University conspiracy.

### 8.21 Serendipity

"I worry about self-archiving because of those lucky happenstances that happen only when browsing index cards, library shelves, and journal contents."

This worry, despite its charm, does not deserve much space: With time, it will become evident that on-screen digital searching and browsing can be every bit as serendipitous as on-paper analog searching and browsing; chance adjacency effects are every bit as potent either way. The searching and browsing will simply be less exhausting to the limbs and fingers.

No rational deterrent to immediate self-archiving in worries about loss of serendipity.

### 8.22 Tenure/Promotion

"I worry about self-archiving because it does not count as refereed publication, and might even interfere with the chances for refereed publication."

Yet another instance of the archiving/publishing conflation: The self-archiving initiative is aimed at freeing *refereed* publication from toll-based access/impact barriers (not from refereeing). Unrefereed preprints do not count as publications on-line any more than they do on-paper.

The other half of this worry is probably a variant of the Copyright (8.10) concerns (q.v.) as well as concerns about Embargo policies (Harnad 2000a, 2000b), both of which are groundless.

No rational deterrent to immediate self-archiving in worries about tenure/promotion.

### 8.23 (your prima-FAQ here ...)

## 9. Related Issues

### 9.1 Napster

Author-end give-aways of their own digital products via self-archiving is the antithesis of consumer-end rip-offs of others' non-give-away digital products via napster (<http://www.napster.com/>) or gnutella (<http://gnutella.wego.com/>).

It is very important to clearly distinguish and distance the two, because any inadvertent or willful conflation of the self-archiving initiative with napster can only retard the progress of the self-archiving initiative toward the optimal and inevitable.

("Information is free" is nonsense: There is and always was both give-away and non-give-away information. Steal the latter and you simply kill the incentive to provide it in the first place.)

### 9.2 Peer-review reform

Peer review is not without its flaws, but improving peer review first requires careful testing of alternative systems, and demonstrating empirically that these alternatives are at least as effective as classical peer review in maintaining the quality of the refereed literature (such as it is). No alternatives have yet been tested or demonstrated effective.

Hence current peer review reform or elimination proposals are merely speculative hypotheses at this time, and red herrings insofar as the freeing of the peer-reviewed literature is concerned: The self-archiving initiative is directed at freeing the current peer-reviewed literature, such as it is, from the impact/access barriers of S/L/P access-tolls, now. It is not directed at freeing the literature from peer review, or at testing or implementing untested alternatives to peer review (Cf. <http://library.caltech.edu/publications/ScholarsForum/042399sharnad.htm> and <http://www.cogsci.soton.ac.uk/~harnad/Ebiomed/com0509.htm#harn45>).

The benefits of freeing the refereed literature now are a sure thing; the benefits (if any) from future alternatives to peer review (if any) are purely hypothetical, and certainly nothing to hold as back from self-archiving to wait for.

### 9.3 "Scholarly skywriting"

An additional benefit of at last having the entire refereed literature online and freed of access/impact barriers is that this "skyreading and skywriting," potentially accelerating the global collaborative, cumulative, self-corrective cycles of human interaction in research almost to the "speed of thought," will help to

increase our planet's collective scholarly and scientific productivity. (Harnad 1990, 1991, 1992, 1995c, Light et al. 2000).

#### 9.4 Embryology of knowledge

An online, interoperable, citation-linked refereed research literature also makes it possible to monitor and measure scholarly/scientific practise, progress and impact in powerful new ways that go far beyond mere citation-impact (e.g. download impact, download immediacy, pre-refereeing impact, user navigational route analysis, revision embryology, searching with online author/paper/journal impact ranking, online co-download/co-citation searching, etc.). (Harnad & Carr 2000).

#### 9.5 Leading horses to the waters of self-archiving vs. getting them to drink

Will the availability of free, interoperable software for creating institution-based Eprint Archives get us to the optimal/inevitable at last? Future historians will have to be the judge (Harnad 1999b). But it is already a historical fact that it is already within reach, and that we have been slow to grasp.

#### APPENDIX A: Eprints.org Software for Creating Institutional and Individual Open Archives

At the 2nd Open Archive Initiative (OAI) meeting in San Antonio in June (<http://www.dlib.org/dlib/june00/06inbrief.html#FOX>) a participant said:

"Open Archiving will not get off the ground until the day I can go to a website, download open-archiving software, then say MAKE ARCHIVE, and an interoperable, OAI-compliant archive is up and running, ready to be filled."

At Southampton, we took this to heart, and applying our experience with the CogPrints archive, designed the generic eprints.org software that fits this bill. A public beta version has been released and has taken over operations at the CogPrints site (<http://cogprints.soton.ac.uk/>). The operational release (December 2000) is free and will be open sourced (and over 100 prospective users worldwide are already signed up).

The eprints.org software is a feature-rich, easily installed, eprint archive system. It runs right "out of the box" with a comprehensive default setup that should serve most institutions' and individuals' needs as it stands. It has also been designed to make it extensively and flexibly re-configurable for customised needs; almost any aspect of the archive's operation can be adapted to suit a particular requirement.



The archive supports the OAI protocol, allowing it to interoperate with other open archives and open archive services, and to be readily upgraded to keep up with OAI revisions.

This adaptability is achieved by using a modular design methodology. The system is divided into two main components: The core archive component, which provides the functionality required for all open archives, and the site-specific component, providing details about exactly what is stored in the archive, how it is presented and how it may be searched. The system is supplied with a richly featured site-specific component that requires minimal changing to set up a fully working, interoperable open archive. When updated revisions of the software become available, the core archive component can be upgraded, and the site retains its identity and data in the site-specific component.

1. The many aspects of the software that can be configured by an institution include:

- The types of record that can be stored in the archive, and what metadata fields to hold with each
- The types of document file (or other data) that can be stored with each record
- The validation checks that are performed on each incoming record, to minimise administrator effort
- The choice of which metadata fields are searchable by users
- The choice of what metadata to present records to the open archives protocol (i.e., specifying how the internal metadata maps to the open archives metadata)
- Full control over the "look and feel" of the archive (in any language: French translations currently being prepared by Helene Bosc and colleagues at <http://inra.fr/>)

2. The software also has the following features:

- "Out of the box" Open Archives Initiative interoperability
- Simple but very powerful depositing interface
- Local browsing and searching features
- Inter- and intra-linking potential (papers, versions, comments, responses)
- Moderation buffer for incoming deposits
- Site maintenance via a WWW interface
- E-Mail subscription service for users

It is simple to add extra functionality to an archive in the site-specific component of the software. This means that the archive can be used by institutions,

individuals, journals or any other organisation wishing to interoperate with Open Archive services.

## APPENDIX B: Some Relevant Chronology and URLs

(See also Peter Suber's fuller timeline at the Free Online Scholarship site: <http://www.earlham.edu/~peters/fos/timeline.htm>.)

- *Psycology* (Refereed On-Line-Only Journal) (1989)  
<http://www.cogsci.soton.ac.uk/psycology>
- "Scholarly Skywriting" (1990)  
<http://cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad90.skywriting.html>
- Physics Archive (1991)  
<http://arxiv.org>
- "PostGutenberg Galaxy" (1991)  
<http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad91.postgutenberg.html>
- "Interactive Publication" (1992)  
<http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad92.interactivpub.html>
- Self-Archiving ("Subversive") Proposal (1994)  
<http://www.arl.org/scomm/subversive/toc.html>
- "Tragic Loss" (Odlyzko) (1995)  
<http://www.research.att.com/~amo/doc/tragic.loss.txt>
- "Last Writes" (Hibbitts) (1996)  
<http://www.law.pitt.edu/hibbitts/lastrev.htm>
- NCSTRL: Networked Computer Science Technical Reference Library (1996)  
<http://cs-tr.cs.cornell.edu>
- University Provosts' Initiative (1997)  
<http://library.caltech.edu/publications/ScholarsForum/>
- *CogPrints*: Cognitive Sciences Archive (1997)  
<http://cogprints.soton.ac.uk>
- *Journal of High Energy Physics* (Refereed On-Line-Only Journal) (1998)  
<http://jhep.cern.ch/>
- *Science* Policy Forum (1998)  
<http://www.sciencemag.org/cgi/content/full/281/5382/1459>
- *American Scientist* Forum (1998)  
<http://amsci-forum.amsci.org/archives/American-Scientist-Open-Access-Forum.html>
- *OpCit*: Open Citation Linking Project (1999)  
<http://opcit.eprints.org>
- *E-biomed*: Varmus (NIH) Proposal (1999)  
<http://www.nih.gov/about/director/pubmedcentral/pubmedcentral.htm>

- Open Archives Initiative (1999)  
<http://www.openarchives.org>
- Cross-Archive Searching Service (2000)  
<http://arc.cs.odu.edu>
- *Eprints*: Free OAI-compliant Eprint-Archive-creating software (2001)  
<http://www.eprints.org>
- FOS: Free Online Scholarship Movement (2001)  
<http://www.earlham.edu/~peters/fos/timeline.htm>
- *BOAI*: Budapest Open Access Initiative (2002)  
<http://www.soros.org/openaccess>
- *Self-Archiving FAQ* (2002)  
<http://www.eprints.org/self-faq/>
- Harnad Home Pages  
<http://cogsci.soton.ac.uk/harnad/>  
<http://www.princeton.edu/~harnad/>

## References

- Carr, L., Hitchcock, S., Hall, W. & Harnad, S. (2000). A usage based analysis of CoRR [A commentary on "CoRR: a Computing Research Repository" by Joseph Y. Halpern]. *ACM SIGDOC Journal of Computer Documentation*, May 2000.  
<http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad00.halpern.htm>  
<http://www.garfield.library.upenn.edu/papers/cseimpactfactor05092000.html>
- Duranceau, E. & Harnad, S. (1999). Electronic Journal Forum: Resetting Our Intuition Pumps for the Online-Only Era: A Conversation With Stevan Harnad. *Serials Review*, 25(1), 109-115.  
<http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad99.ejforum.html>
- Garfield, E. (1955). Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas. *Science*, 122, 108-111  
[http://www.garfield.library.upenn.edu/papers/science\\_v122\(3159\)p108y1955.html](http://www.garfield.library.upenn.edu/papers/science_v122(3159)p108y1955.html)
- Harnad, S. (1990). Scholarly Skywriting and the Prepublication Continuum of Scientific Inquiry. *Psychological Science*, 1, 342 - 343 (reprinted in *Current Contents*, 45, 9-13, November 11 1991).  
<http://cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad90.skywriting.html>
- Harnad, S. (1991). Post-Gutenberg Galaxy: The Fourth Revolution in the Means of Production of Knowledge. *Public-Access Computer Systems Review*, 2(1), 39-53 (also reprinted in *PACS Annual Review*, 2, 1992; and in R. D. Mason, Ed., *Computer Conferencing: The Last Word*. Beach Holme Publishers, 1992; and in M. Strangelove & D. Kovacs: *Directory of Electronic Journals, Newsletters, and Academic Discussion Lists*, A. Okerson, Ed., 2nd edition. Washington, DC, Association of Research Libraries, Office of Scientific & Academic Publishing, 1992); and in Hungarian translation in *REP-*

- LIKA, 1994; and in Japanese in *Research and Development of Scholarly Information Dissemination Systems*, 1994-1995.  
<http://cogsci.soton.ac.uk/harnad/Papers/Harnad/harnad91.postgutenberg.html>
- Harnad, S. (1992). Interactive Publication: Extending American Physical Society's Discipline-Specific Model for Electronic Publishing. *Serials Review*, Special Issue on Economics Models for Electronic Publishing, 58-61.  
<http://cogsci.soton.ac.uk/harnad/Papers/Harnad/harnad92.interactivpub.html>
- Harnad, S. (1994). A Subversive Proposal. In Ann Okerson & James O'Donnell (Eds.), *Scholarly Journals at the Crossroads: A Subversive Proposal for Electronic Publishing*. Washington, DC.: Association of Research Libraries.  
<http://www.arl.org/scomm/subversive/toc.html>
- Harnad, S. (1995c). Interactive Cognition: Exploring the Potential of Electronic Quote/Commenting. In B. Gorayska & J. L. Mey (Eds.), *Cognitive Technology: In Search of a Humane Interface* (pp. 397-414). Elsevier P.  
<http://cogsci.soton.ac.uk/harnad/Papers/Harnad/harnad95.interactive.cognition.html>
- Harnad, Stevan (2003) Back to the Oral Tradition Through Skywriting at the Speed of Thought. *Interdisciplines*.  
<http://www.interdisciplines.org/defispublicationweb/papers/6>
- Harnad, S. (1998b). On-Line Journals and Financial Fire-Walls. *Nature*, 395(6698), 127-128. <http://www.cogsci.soton.ac.uk/~harnad/nature.html>
- Harnad, S. (1998/2000). The invisible hand of peer review. *Nature* [online] (5 Nov. 1998) <http://helix.nature.com/webmatters/invisible/invisible.html>  
 Longer version in *Exploit Interactive* 5 (2000):  
<http://www.exploit-lib.org/issue5/peer-review/>  
<http://www.cogsci.soton.ac.uk/~harnad/nature2.html>
- Harnad, S. (1999a). The Future of Scholarly Skywriting. In A. Scammell (Ed.), *I in the Sky: Visions of the information future*. Aslib, November 1999.  
<http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad99.aslib.html>
- Harnad, S. (1999b). Free at Last: The Future of Peer-Reviewed Journals. *D-Lib Magazine*, 5(12) December 1999.  
<http://www.dlib.org/dlib/december99/12harnad.html>
- Harnad, S. (1999c). Advancing Science By Self-Archiving Refereed Research. *Science dEbates* [online] 31 July 1999.  
<http://www.sciencemag.org/cgi/eletters/285/5425/197#EL12>
- Harnad, S. (2000a). E-Knowledge: Freeing the Refereed Journal Corpus Online. *Computer Law & Security Report*, 16(2) 78-87 (rebuttal to Bloom Editorial in Science and Relman Editorial in New England Journal of Medicine).  
<http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad00.scinejm.htm>
- Harnad, S. (2000b). Ingelfinger Over-Ruled: The Role of the Web in the Future of Refereed Medical Journal Publishing. *The Lancet Perspectives*, 256, (December Supplement).  
<http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad00.lancet.htm>
- Harnad, S. (2001a). AAAS's Response: Too Little, Too Late. *Science dEbates* [online] 2 April 2001.

- <http://www.sciencemag.org/cgi/eletters/291/5512/2318b>  
 Fuller version: <http://www.cogsci.soton.ac.uk/~harnad/Tp/science2.htm>
- Harnad, S. (2001b). The Self-Archiving Initiative. *Nature*, 410, 1024-1025.  
<http://www.cogsci.soton.ac.uk/~harnad/Tp/nature4.htm>  
*Nature WebDebates* version: <http://www.nature.com/nature/debates/e-access/Articles/harnad.html>  
 Fuller version: <http://www.cogsci.soton.ac.uk/~harnad/Tp/selfarch.htm>
- Harnad, S. (2001c). The Self-Archiving Alternative. *Nature WebDebates*.  
<http://www.nature.com/nature/debates/e-access/index.html>  
<http://www.cogsci.soton.ac.uk/~harnad/Tp/nature3.htm>
- Harnad, S. (2001d). The (Refereed) Literature-Liberation Movement. *New Scientist*. <http://www.cogsci.soton.ac.uk/~harnad/Temp/newsci1.htm>  
 Fuller version: <http://www.cogsci.soton.ac.uk/~harnad/Temp/newscientist.htm>
- Harnad, S. (2001e). Research Access, Impact and Assessment. *Times Higher Education Supplement*. <http://www.cogsci.soton.ac.uk/~harnad/Tp/thes1.html>
- Harnad, S. & Carr, L. (2000). Integrating, Navigating and Analyzing Eprint Archives Through Open Citation Linking (the OpCit Project). *Current Science*. 79(5), 629-638.  
<http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad00.citation.htm>  
<http://www.iisc.ernet.in/~currsci/sep102000/629.pdf>
- Harnad, S. & Hemus, M. (1997). All or None: there Are No Stable Hybrid or Half-Way Solutions for Launching the Learned Periodical Literature in the PostGutenberg Galaxy. In I. Butterworth (Ed.), *The Impact of Electronic Publishing on the Academic Community*. London: Portland Press.  
<http://cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad97.hybrid.pub.html>
- Harnad, S., Carr, L. & Brody, T. (2001). How and Why To Free All Refereed Research From Access- and Impact-Barriers Online, Now.  
<http://www.cogsci.soton.ac.uk/~harnad/Tp/science.htm>
- Harnad, S., Carr, L., Brody, T. & Oppenheim, C. (2003) Mandated online RAE CVs Linked to University Eprint Archives: Improving the UK Research Assessment Exercise whilst making it cheaper and easier. *Ariadne* 35 (April 2003).  
<http://www.ariadne.ac.uk/issue35/harnad/>
- Harnad, S., Varian, H. & Parks, R. (2000). Academic publishing in the online era: What Will Be For-Fee And What Will Be For-Free? *Culture Machine* 2 (Online Journal). <http://www.cogsci.soton.ac.uk/~harnad/Temp/Varian/new1.htm>  
[http://culturemachine.tees.ac.uk/frm\\_fl.htm](http://culturemachine.tees.ac.uk/frm_fl.htm)
- Hayes, P., Harnad, S., Perlis, D. & Block, N. (1992) Virtual Symposium on Virtual Mind. *Minds and Machines*, 2(3) 217-238.  
<http://cogsci.soton.ac.uk/harnad/Papers/Harnad/harnad92.virtualmind.html>
- Hitchcock, S. Carr, L., Jiao, Z., Bergmark, D., Hall, W., Lagoze, C. & Harnad, S. (2000). Developing services for open eprint archives: globalisation, integration and the impact of links. *Proceedings of the 5th ACM Conference on Digital Libraries*. San Antonio Texas June 2000.  
<http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad00.acm.htm>
- Lawrence, S. (2001a) Online or Invisible? *Nature*, 411(6837), 521.

- <http://www.neci.nec.com/~lawrence/papers/online-nature01/>
- Lawrence, S. (2001b) Free online availability substantially increases a paper's impact. *Nature Web Debates*.  
<http://www.nature.com/nature/debates/e-access/Articles/lawrence.html>
- Light, P., Light, V., Nesbitt, E. & Harnad, S. (2000). Up for Debate: CMC as a support for course related discussion in a campus university setting. In R. Joiner (Ed.), *Rethinking Collaborative Learning*. London: Routledge (in press).  
<http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad00.skyteaching.html>
- Odlyzko, A.M. (1998). The economics of electronic journals. In R. Ekman & R. Quandt (Eds.), *Technology and Scholarly Communication*. Univ. Calif. Press.  
<http://www.dtc.umn.edu/~odlyzko/doc/eworld.html>
- Odlyzko, A.M. (1999a). Competition and cooperation: Libraries and publishers in the transition to electronic scholarly journals. *Journal of Electronic Publishing*, 4(4) (and in *J. Scholarly Publishing*, 30(4), July 1999, 163-185; the definitive version to appear in *The Transition from Paper: A Vision of Scientific Communication in 2020*, S. Berry and A. Moffat, Eds., Springer, 2000.  
<http://www.press.umich.edu/jep/04-04/odlyzko0404.html>  
<http://www.research.att.com/~amo/doc/competition.cooperation.pdf>
- Odlyzko, A.M. (2002). The rapid evolution of scholarly communication." *Learned Publishing*, 15, 7-19.  
<http://rosina.catchword.com/vl=11319436/cl=24/fm=docpdf/nw=1/rpsv/catcword/alpsp/09531513/v15n1/s2/p7>  
<http://www.si.umich.edu/PEAK-2000/odlyzko.pdf>
- Oppenheim, C. (2001). The legal and regulatory environment for electronic information. *Infonortics*.  
<http://www.infonortics.com/publications/legal4.html>