

## **Archive policy for the Cornell Computer Science Department**

On December 1, 2004, the Cornell Computer Science Faculty by consensus adopted the following policy, which was proposed by Bill Arms, Joe Halpern and Steve Vavasis.

All papers emanating from the department will be saved in a publicly accessible archive like [arxiv.org](http://arxiv.org).

This document will attempt to explain the rationale and implementation of the policy.

### **Rationale for the policy**

A crisis has been evolving in the past few years in the realm of scholarly publishing because commercial journals have raised their prices substantially without a proportional benefit to the community of authors or readers. For example, the EMPS (Engineering, Math and Physical Sciences) library at Cornell has seen a 9% subscription increase in just the past year. The worst offender seems to be Elsevier, which publishes many CS journals.

A second looming concern with scholarly publishing is that commercial publishers are using pricing policies to push libraries into switching to all-electronic subscription. All-electronic subscription gives the commercial publisher unprecedented control over who can read articles and for what purposes those articles are used. Furthermore, an electronic subscription means that the publisher expands its role to become also the archivist of the material. There is no reason to believe that a company like Elsevier is qualified to usurp the role traditionally filled by libraries as the archivist of scholarly work over a period of decades or centuries. For more information about the problems faced by university libraries, please visit the [home page](#) of the SPARC project of the Association of Research Libraries.

An obvious solution to these problems is for the academic community as a whole to create its own archive under the control of scholars rather than a corporate board of directors. This is the goal behind arxiv.org. We believe that all academics ought to include their publications in this kind of archive. Therefore, we are establishing this as a departmental policy. We would like to establish it as a policy for the whole world, but we have to start somewhere!

Naturally, a member of the department could easily follow this policy on his or her own initiative without the existence of a departmentwide policy. Indeed, several of us already archive our papers as a matter of course because archiving brings several benefits to the author including enhanced visibility of the result and proof of precedence of discovery. But we believe there are three reasons why it is useful to make archiving an official policy of the department.

1. By making it a policy, we are making a public statement in favor of open archiving.
2. There is clearly a snowball effect at work: the more computer scientists who archive, the more useful the archive becomes, and hence more people will archive, etc.

3. If archiving becomes a policy, then the University Library, which has considerable expertise in the copyright issues involved, can help us to make sure that we protect our right to archive and distribute our materials when we sign journal copyright transfer agreements.

## **FAQ concerning this policy**

### **What is archiving?**

Archiving means storing and managing a document in a way that will ensure its availability over a long period of time, e.g., decades or centuries. University libraries are usually considered archival repositories. In contrast, personal home-pages are generally not considered archival repositories.

### **What makes archiving special?**

In the past, archiving has meant that the document should be printed with long-lasting inks on long-lasting papers and stored in a stable environment (away from direct sunlight, high humidity, etc.) In addition, archiving also implies good cataloging mechanisms to make sure documents can be found when they are sought. In the internet era, archiving means that documents should be stored in a stable and well-backed up medium. It also means that a group of archivists must over the years take responsibility for updating digital documents in the case that encoding standards (such as PDF) and retrieval protocols (such as http) evolve.

### **What is [arxiv.org](http://arxiv.org)?**

Arxiv.org is a web-based archival repository for scientific documents. Currently, it has three subject areas: physics, mathematics and computer science. It is supported by the Cornell University library and contains hundreds of thousands of papers from around the world. It is run by a self-perpetuating committee of academics. It was founded by Paul Ginsparg while he was at Los Alamos National Laboratory. Paul is currently a Professor of Physics and of Computing and Information Science at Cornell.

### **Why is [arxiv.org](http://arxiv.org) considered archival?**

- First, [arxiv.org](http://arxiv.org) has a policy that once a paper is submitted, it cannot be removed. (But it can be updated, and errata can be published, etc.) This is similar in spirit to the idea that a paper, once published in a journal, can never be "depublished".
- Second, [arxiv.org](http://arxiv.org) has implemented policies to try to improve the chances that a paper archived today will still be accessible in 100 years. For example, [arxiv.org](http://arxiv.org) requests that papers written in Latex be submitted in their Latex form (rather than in postscript or PDF) because Latex is a fairly stable, human-readable and well-documented representation of documents. Furthermore, Latex documents contain more information than the PDF output from Latex, so they can be catalogued more reliably.

- Third, arxiv.org will remain the responsibility of the Cornell University Library. As one of its primary mandates from the University, the Library is accountable for ensuring that designated information in all formats, including the content of the arxiv, is maintained and accessible for the long term.

### **Why should I want to submit my paper to arxiv.org?**

Please see the [rationale section](#) of this document for some reasons why archiving your paper is beneficial to your career.

### **But what if I want to keep my Latex source confidential?**

There are several answers to this question. When you submit your Latex paper, you can check a box indicating that the Latex should not be distributed. In addition, if your paper has comments in it that you would prefer to keep confidential, you can run a perl script (available on the arxiv website) that strips comments prior to submission.

### **But many journals, e.g., ACM and SIAM, think that PDF is good enough. Really, why can't arxiv.org just use PDF?**

The maintainers of arxiv.org have found incompatibilities in versions of PDF that may render your document unreadable by scholars over a very long period of time. There is currently a proposal for an archival version of pdf called [PDF/A](#). If this proposal becomes reality, then arxiv.org will probably allow PDF/A submission.

### **What if my document is in Word?**

In this case, arxiv allows you to submit the PDF version of your document.

### **Doesn't archiving violate a journal's copyright policy?**

First, note that you can alter copyright transfer agreements to preserve more rights for yourself. Naturally, a journal might reject the paper if it disagrees with your alterations to the agreement, but we have heard that many people have successfully altered these agreements without adverse consequences. Later, we will post some possible alterations that people have successfully used on copyright transfer agreements.

Assuming you don't alter the agreement, you are subject to the terms of it. Here are the policies of some of the larger CS publishers.

- **ACM.** The [relevant ACM policy](#) states that, prior to acceptance you can post the preprint version of the paper anywhere, but that after acceptance you should add an ACM copyright note to the posted version. After acceptance, you can post the accepted paper on your home page but not on a server like arxiv.org unless you first obtain ACM permission. In the case of Cornell University employees, an author-prepared version of the paper can be posted to arxiv.org even after

acceptance because ACM allows posting on a "publicly accessible server of their employer". For Cornell employees, arxiv.org counts as an employer's server.

There is a potential difficulty with ACM conferences that use blind reviewing. See further remarks below.

- **SIAM.** The SIAM [copyright transfer agreement](#) allows you to post the preprint version anywhere including arxiv.org, and allows you to post the final version on your personal home page but not on a server like arxiv.org.
- **IEEE.** The [relevant IEEE policy](#) states that you can post the preprint version of the paper anywhere including arxiv.org, but, upon acceptance, you are required to replace it with the accepted version that includes the IEEE copyright notice.
- **Elsevier.** The [relevant Elsevier policy](#) states that you can post a preprint version of the paper anywhere including arxiv.org, and you can post the final version on your home page (but not arxiv.org). You are allowed to post revised versions made during the refereeing process on an employer website. For Cornell University employees, this includes arxiv.org.
- **Springer.** The [relevant Springer policy](#) states that you can post a preprint version of the paper anywhere including arxiv.org but not the final version. Springer also has a program called [Open Choice](#) in which you pay Springer \$3000, and in return, they will make the final PDF version of your paper available on their server for free for anyone with web access.
- **Wiley.** I have not been able to figure out the [Wiley copyright transfer policy](#). I sent email to Wiley in January, 2005 to request clarification but have not yet received a definitive answer.

### What about embargo policies?

Some journals have embargo policies stating that a result may not be disclosed in any form prior to journal submission. Some very well known journals like *Science* have a firm embargo policy. An example of a CS publication with an embargo policy is SIGGRAPH proceedings. SIGGRAPH uses a blind reviewing system (i.e., paper reviewers are not told the names of the authors), and any web-distribution of a paper would undermine the possibility that it could be blindly reviewed. We are currently investigating specifically whether there is a workaround for SIGGRAPH. In the case of Cornell University employees, the matter appears to be moot since, as mentioned above, ACM gives authors permission to post papers after acceptance on their employers' websites.

Naturally, the journal's policy overrides this policy, i.e., we are not suggesting that anyone should violate a journal's policy in order to follow this one. On the other hand, if you are a believer in archiving and regularly submit papers to journals and conferences with embargo policies, then you can use this document as an argument to convince the journal to loosen its embargo policies.

**Does arxiv.org undermine the traditional refereeing process? In other words, what will happen if everyone starts submitting papers to arxiv.org and they cease submitting them to refereed journals?**

This is an interesting question that will need to be periodically revisited. The experience so far in physics (which has the largest subject area of arxiv.org) is that the refereeing process has not been abandoned, i.e., the arxiv.org papers usually end up in refereed journals as well.

**What if I simply don't want to archive my papers?**

Compliance with this new policy of archiving papers is entirely voluntary. On the other hand, if you choose to ignore the policy, give some consideration to the reasons for this choice. If there are technical reasons not already covered by this FAQ, please bring them to the attention of an arxiv.org board members, for example, Joe Halpern.

**Acknowledgements**

I received helpful comments on this writeup from Bill Arms, Joe Halpern and Ross Atkinson.

*Stephen A. Vavasis, Department of Computer Science, 4130 Upson Hall, Cornell University, Ithaca, NY 14853, vavasis@cs.cornell.edu. Last update: April 12, 2005.*