From Preprints to E-prints:

The Rise of Electronic Preprint Servers in Mathematics

Allyn Jackson

One of the earliest preprint servers in mathematics was launched in July 1991 in the mathematics department of the University of Texas at Austin. Called mp_arc, for "mathematical physics preprint archive", it was the brainchild of Hans Koch, Rafael de la Llave, and Charles Radin, who had no grander scheme than to provide themselves and their colleagues in mathematical physics with an efficient and organized way of exchanging preprints. In the state next door, the hep_th preprint server went online in August that same year at Los Alamos National Laboratories (LANL) in New Mexico. Founded by physicist Paul Ginsparg, hep_th, which stands for "high energy physics—theoretical", housed preprints in that rapidly moving field. Both mp_arc and hep_th operated by e-mail and ftp; the World Wide Web had not yet been invented.

These two servers made the jump to the Web and still exist today, and each has been successful in its own way. But their trajectories have been very different. The mp_arc has remained a small archive serving mathematical physicists, with about 2,700 papers and about 850 subscribers to its e-mail-notification service. By contrast, hep_th has grown into a behemoth now known as the arXiv. Its size is a moving target: At the time this article was written, the arXiv contained over 170,000 papers in all areas of physics, as well as mathematics, nonlinear science, and computer science. Around 25,000 people get daily e-mail about new postings to the arXiv, and around 35,000 users access the arXiv every day.

Preprint servers are fast becoming an integral part of the research culture in mathematics. Posting papers on preprint servers is now "part of the publication process", says Dale Alspach of Oklahoma State University, who runs the Banach Space Archive. Over the approximately ten years that preprint servers have been in existence, their holdings have grown in size and in value. Attempts to manage the evolution of this new tool have produced considerable ferment within the worldwide mathematical community. Preprint servers have also raised tough questions about long-term electronic storage, copyrights, and peer review.

From Preprints to E-prints

In the days before the Internet, preprints were on paper and were usually circulated by postal mail or handed out at lectures. Preprints are now circulated mostly electronically, often through e-mail but increasingly through websites. A preprint server is an automated electronic mechanism, usually Web-based, for exchanging preprints of scholarly articles. The servers are fully accessible to anyone. Authors post their preprints; readers retrieve the preprints they are interested in. There are no gatekeepers judging the quality of the posted material (though oversight is exercised to eliminate inappropriate postings), and there are no access fees.

Sometimes preprints are removed from a server after publication, but because many servers retain the material in perpetuity, they call themselves "preprint archives". What is more, a few journals

Allyn Jackson is senior writer and deputy editor of the Notices. *Her e-mail address is* axj@ams.org.

have begun to post published articles on these servers, rendering the term "preprint" a misnomer. Thus the phrase "e-print archive" has come into use. "People sometimes say 'e-print' instead of 'preprint'," says Greg Kuperberg of the University of California at Davis, who has worked extensively on the mathematics section of the arXiv. "I go further. I say 'article' to make the same point." Indeed, the notion that this material is tentative or ephemeral is disappearing. Some believe that such archives will evolve to become the custodians of the primary research literature.

Such an evolution already seems to be occurring in physics, due to the arXiv. Some physicists who no longer feel the need to add to their publication lists have posted influential papers on the arXiv with no intention of submitting those papers to journals. More and more often in physics (and to a lesser extent in mathematics) one sees arXiv numbers given as references to papers cited in bibliographies. In the past, when research articles appeared only on paper, the number of times an average article was cited grew slowly over a period of years. Citation analyses have shown that, for physics papers, the arXiv has shortened this citation time-lag to just months.

Subject-based Servers: Small, Efficient, Folksy

Worldwide, there are hundreds of preprint servers in mathematics, and they can be divided into two major categories: subject-based servers and general servers. The Directory of Preprint and e-Print Servers, provided by the AMS on its website, lists seventeen subject-based servers, and these together contain 6,000 to 7,000 papers. These servers are typically run by mathematicians themselves and provide easy-to-use, low-tech features for small communities of dedicated users. The mp_arc, which is limited to the subject of mathematical physics, is an example of a subject-based archive. Koch says that the mp_arc shut down its old gopher server after nobody had accessed it in a year. Although most users now access the mp_arc through the Web, the archive maintains an e-mail interface, because a small percentage of papers is still posted that way. "We keep it at a low technical level as far as access is concerned, because people in different parts of the world don't have all the latest gimmicks," he notes.

These servers put a premium on getting the user to the desired material as quickly as possible. For example, the home page of the K-theory Preprint Archives serves up the full list of about 500 papers, giving titles and authors. From there one click retrieves the abstract of a paper, and another click downloads a `.dvi` or PostScript file containing the full paper. Some servers offer a bigger range of file formats, including TEX, LATEX, and PDF (Portable Document Format). A search function is a common feature: On some servers, like mp_arc, the full text of the papers is indexed and searchable, but on most others one can search only the "metadata"—that is, the titles, author names, and abstracts. Despite the lack of ornamentation, many of these sites have distinctive personalities. The Hopf Topology Archive, whose e-mail notification service reaches about 400 people, has a downright folksy feel. On the home page of this archive, its founder, Clarence Wilkerson of Purdue University, tells users he recently found an old roll of film with pictures from a conference in the early 1980s. "Write me if you can figure out when and where this was," he requests. "Current thought is the UWO conference from 1981." He is clearly addressing a small clan.

Why do mathematicians start preprint servers? "It wasn't a grand vision," explains Wilkerson, who started the Hopf Topology Archive in 1992. He wanted to avoid the expense and delays of sending paper preprints through the mail, especially overseas, and he also hoped to publicize his and his colleagues' work. With about 700 papers, the archive has clearly developed into a useful resource. Wilkerson points to some unforeseen side benefits as well. "The archive has allowed me as a topologist to get a view of what's currently going on in the field," he notes. The archive has also attracted authors from distant corners of the globe, who do not have easy access to paper versions of preprints. Says Wilkerson, "A psychological payback is that I often hear profuse thanks for having this service available."

General Servers: A Variety of Flavors

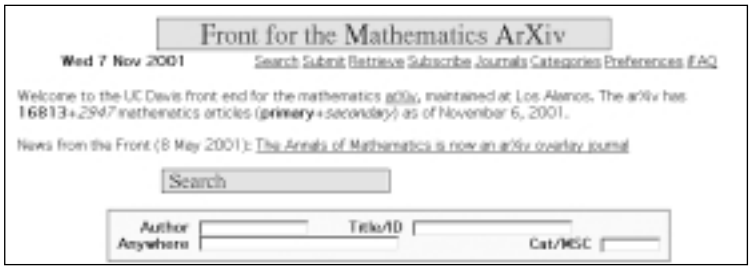
The second category of preprint servers, that of general servers, includes those based at mathematics institutes or in mathematics departments. Most institutes have a place where visitors can post preprints of articles they worked on during their visits, and many mathematics departments have central servers where faculty can deposit their papers. There are hundreds of such servers, and they vary greatly in size, ranging from, for example, the tiny server at the mathematics department of the Royal Swedish Institute of Technology, which has about sixty papers, to the server at the Institute for Mathematics and its Applications at the

University of Minnesota, which has about 1,800. According to Kuperberg, mathematics institute and departmental servers together contain roughly as many papers as the mathematics section of the arXiv (which as of this writing had 16,400 papers).

A preprint server can be an important part of the image that an institute or department projects on the World Wide Web: A moribund departmental server, for example, can leave the impression that no one in the department is doing any research. When the Erwin Schrödinger Institute (ESI) was launched in Vienna in 1993, ESI immediately set up a preprint server, which was one of the very first to use HTML (Hypertext Markup Language). Today the server contains about 1,000 preprints. "We need the preprint server as documentation of our productivity," explains ESI director Peter Michor. The institute is funded by the government of Austria, which like all funders wants proof that its money was well spent.

Another example of a general preprint server is the mathematics section of the arXiv (pronounced the same as "archive"; the X is pronounced like the X in TEX). The arXiv differs from institute or departmental servers in not being closely tied to its home institution, Los Alamos National Laboratories (LANL). In fact, this year Paul Ginsparg accepted an offer to join the new Faculty of Computing and Information at Cornell University (with a joint position in the physics and computer science departments), and the main arXiv site will move with him. Cornell has become in recent years a center for innovation in electronic communications for academia, making it a natural home for the arXiv. Another difference is that the arXiv has outside funding: For the past few years, its funding totaled about $300,000 per year from the Department of Energy and the National Science Foundation. Most of the funding goes toward the salaries of the three staff members, including Ginsparg, who develop and maintain the arXiv; the hardware costs are comparatively minor. The Department of Energy funding will end with the move to Cornell, while the National Science Foundation will continue to fund research and development supporting the arXiv. The Cornell University Library will cover the arXiv's basic infrastructure and maintenance as a "special collection" of the library.

The dedicated resources and attention put into the arXiv over the years have resulted in a highly robust system. The robustness is partly due to refinements in the software; the arXiv is, for example, one of the few preprint servers that has a TEX "autocompiler" that figures out what version of TEX a posted paper is in and then automatically compiles the paper in the appropriate version. In addition, seventeen mirror sites spread across six continents ensure that the arXiv data is preserved and always accessible. But, says Greg Kuperberg, "What I think sets the arXiv apart [from other preprint servers] is its oversight even more than its software. It has a full-time paid staff, several dedicated volunteer helpers, and an array of moderators and advisors. This escalation of policy and structure is just what you would expect for a system that now gets 30,000 submissions a year."



The arXiv has gained a reputation for being, as an article in the online magazine *Searcher* put it, "somewhat user-unfriendly". However, the site's documentation, updated and refined over a decade, is clear and ample. First-time users in mathematics might find it easier to get started at the Front for the Mathematics arXiv, created by Kuperberg. The Front is an "overlay" of the arXiv, which means that it provides a different user interface for accessing the arXiv holdings. With more graphics and a more polished feel, the Front offers the same functionality as the arXiv. One additional feature available on the Front but not on the arXiv is an alphabetical listing of authors in mathematics; from this list, one can click on an author's name and retrieve the full list of that author's papers that have been posted on the arXiv.

Another kind of general preprint server is the "umbrella server", which does not contain actual papers but instead provides links to where the papers reside. The prime example of an umbrella server in mathematics is MPRESS (Mathematical Preprint Search System). MPRESS grew out of D-MathNet, started in 1995 by the Deutsche Mathematiker Vereinigung (German Mathematical Society). D-MathNet provides a centralized way to access papers on about 40 mathematics department preprint servers in Germany. Created by Judith Plümer and Roland Schwänzl of the Universität Osnabrück, D-MathNet gives the departments an easy way of organizing preprint metadata into a form that can be "harvested" by Web-crawling robots. In 1998, the European Mathematical Society launched a project to expand the coverage of D-MathNet to servers outside Germany, and thus MPRESS was born. The Committee on Electronic Information and Communication of the International Mathematical Union has now appointed a committee to support the further development of MPRESS. MPRESS draws on not only the servers

Surfing for Preprints

Below is a listing of the URLs of preprint servers mentioned in this article. A comprehensive listing of preprint servers in mathematics may be found in the AMS Directory of Preprint and e-Print Servers, on the Web at http://www.ams.org/global-preprints/.

arXiv
http://arXiv.org

Cite-Base Search
http://cite-base.ecs.soton.ac.uk/help/index.php3

Chemistry Preprint Server (sponsored by Elsevier Science)
http://preprint.chemweb.com

CogPrints
http://research.ecs.soton.ac.uk/projects/CogPrints.html

The Front for the Mathematics arXiv
http://front.math.ucdavis.edu

Hopf Topology Archive
http://hopf.math.purdue.edu/pub/hopf.html

Institute for Mathematics and its Applications, University of Minnesota
http://www.ima.umn.edu/preprints/new.preprintlist.html

K-theory Preprint Archive
http://www.math.uiuc.edu/K-theory

Mathematics Preprint Server (sponsored by Elsevier Science)
http://wwww.mathpreprints.com

mp_arc
http://www.ma.utexas.edu/mp_arc

MPRESS
http://mathnet.preprints.org

ResearchIndex (formerly CiteSeer)
http://citeseer.nj.nec.com

Royal Swedish Institute of Technology
http://www.math.kth.se/math/harvest/brokers/tritamat

Schrödinger Institute preprint archive
http://www.esi.ac.at/ESI-Preprints.html

under D-MathNet, but also servers in France and Austria, as well as the arXiv and a few subject-based servers in the United States; there are plans to bring yet more servers within the reach of MPRESS. A single search command typed into MPRESS allows one to search the metadata of around 40,000 papers (this figure includes duplicate papers residing in more than one server). The usefulness of MPRESS is clear: Plümer says that in 1999 some 50,000 queries were typed into the MPRESS search engine.

Last year saw the establishment of a new kind of preprint server: In May 2001 Elsevier Science launched the first mathematics preprint server sponsored by a commercial publisher (Elsevier started a preprint server in chemistry in 2000 and is considering starting one in computer science). Neither the server's URL, nor the site itself, advertises the Elsevier sponsorship. The server operates free of charge, and Elsevier has pledged to keep it free always. Anyone may post papers to the server, whether or not those papers are destined to be submitted to Elsevier journals. The Elsevier server has some unusual features: For example, users can rate papers on a scale of zero to five stars and can retrieve a list showing the papers in order from the highest ranked to the lowest. Michiel Kolman, Elsevier's publishing director for mathematics and computer science, says that authors publishing in journals on which Elsevier holds the copyright will be given permission to post the final versions of their papers on the Elsevier preprint server, making the papers available for free. Kolman says that the preprint server will be one component of the "math portal" that Elsevier hopes to introduce in 2002. This portal will provide a search service for metadata of articles published by Elsevier and other publishers. The basic idea of establishing these free services is to create Web traffic toward ScienceDirect, Elsevier's online service for subscribers to its journals.

What Other Sciences Are Doing

In physics, the arXiv has become an indispensable mode of scientific exchange. Apart from physics, mathematics is the field in the natural sciences that makes heaviest use of preprint servers. But use of these servers is not uniform across mathematics. For example, mathematical areas with ties to string theory, such as algebraic and differential geometry, are among the biggest users of preprint servers, due to the influence of the arXiv: During the growth of string theory in the 1990s, the arXiv became the central clearinghouse for papers in that rapidly moving field. The use of preprint servers is lower in, for example, numerical analysis, in part because this area has long-established traditions of using electronic tools like e-mail newsgroups and mailing lists for exchanging information about the latest developments.

Like numerical analysis, computer science lacks a tradition of using preprint servers. Rather, the tradition has been to post one's papers on one's own home page or on a departmental server. This fragmented system got a unifying boost when a service called CiteSeer was established by the NEC Research Institute, a computer science think tank

in Princeton, New Jersey. ResearchIndex, as CiteSeer is now called, trolls the home pages of computer scientists on a regular basis, collecting and caching papers. ResearchIndex allows direct postings by authors, and there is now a computer science section of the arXiv, called Computing Research Repository, or CoRR. Joseph Halpern, a computer scientist at Cornell University who chairs the advisory committee for CoRR, notes that, while some computer scientists put their papers on CoRR (and the numbers are increasing every year), the majority seem content to post papers only on their home pages and let ResearchIndex collect them. Papers posted on CoRR will remain in perpetuity, while it is unclear how long the material cached in ResearchIndex will remain there.

In chemistry, the main preprint server is the one sponsored by Elsevier, which was started in August 2000. But with only about 300 preprints submitted since then, the server's growth has been sluggish. Because patents and other proprietary information are often discussed in chemistry papers, the field does not have a well-established culture of exchanging preprints. However, the biggest barriers against preprint servers are found not in chemistry but in the biomedical sciences. Many biomedical journals not only prohibit their authors from posting preprints but also impose embargo rules whereby authors cannot even discuss the contents of an article before it appears in print. In recent years, researchers have found these kinds of rules increasingly restrictive. In 1999, the *British Medical Journal* started NetPrints.org, a repository for preprints of articles in the biomedical sciences. Many of the top biomedical journals—including the *Journal of the American Medical Association*, the *New England Journal of Medicine*, and *Science*—refuse to allow their authors to post preprints on the Web, thereby limiting the usefulness of sites like NetPrints.org.

Embargo rules and patent considerations aside, there is another reason it has been easier to set up preprint servers in mathematics, computer science, and physics than in chemistry and biomedicine: the pervasive use of TEX. David Morrison of Duke University, who chairs the advisory board for the mathematics section of the arXiv, notes that mathematicians inherited the "open source spirit" of TEX from its founder, computer scientist Donald Knuth, who made this revolutionary typesetting language freely available. With TEX, mathematicians have an easy way to exchange files that are independent of users' computer platforms and are convertible into other formats, such as PostScript and PDF. In some areas of science, the use of specialized commercial software packages makes it much more difficult to share files. (Morrison also points out another, unexpected advantage of TEX: "It provides a barrier to participation by amateurs." In other words, because would-be angle-trisectors are unlikely to have mastered TEX, they are discouraged from attempting to post material to preprint servers.)

As the use of preprint servers has spread, a few scientist-led initiatives have sprung up to facilitate the growth of what has been to date mostly a cottage industry. One is the Open Archives Initiative (OAI), of which Ginsparg is a founder. The OAI is an international effort to develop interoperability standards for disseminating content over the Web. In January 2001, the OAI made available protocols for metadata that could facilitate the sharing of papers among different archives. A related effort is the "Self-Archiving Initiative" started by Stevan Harnad, a cognitive scientist at the University of Southampton in the United Kingdom. He has created free archiving software that complies with the OAI standards and that authors can use to construct their own archives. The software creates data that could be "harvested" by, as Harnad calls it, "a global virtual archive". Harnad has begun such an archive, called Cite-Base, which draws its data from the arXiv and from Cogprints, the electronic archive Harnad created in cognitive science.

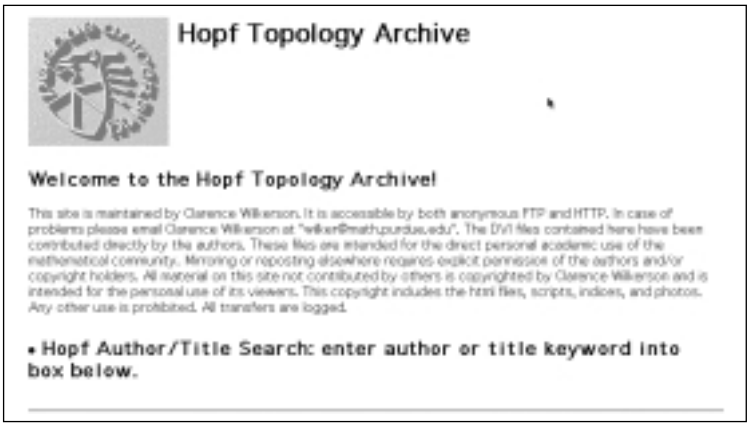
Launching a Server: A Risky Proposition?

Most of the mathematics preprint servers that have thrived have served relatively small communities of mathematicians, numbering in the hundreds, who share common interests. A high level of research activity can also spur a server's growth, as string theory did for the arXiv. However, as Greg Kuperberg points out, starting a preprint server is risky. "Most preprint servers and archives fail, just like startup businesses," he points out. "Most of them either never get off the ground or eventually stall." Some servers had widespread support but died anyway. One such example is the AMS preprint server, which was launched in 1995 and shut down in early 1999.

The idea for starting the AMS preprint server came out of discussions of the Society's Committee on Publications in the early 1990s. The original idea was to run the server using Ginsparg's software from Los Alamos; Ginsparg agreed to share it with the AMS. Eventually this plan was scrapped because of concerns about whether the AMS could properly develop and maintain the software. Ginsparg now says that, with hindsight, the AMS probably made the right decision, because the advent of the World Wide Web necessitated major design changes in the Los Alamos archive. In the end the AMS decided to develop the software on its own. Surrounding the effort was a great deal of uncertainty about what impact emerging technologies like preprint servers might have on traditional journals. Richard Palais of Brandeis

University, who was chair of the Committee on Publications at the time, recalls a question that kept arising: "If people can get the papers electronically before publication, will they still buy the paper journals anymore?"

The AMS preprint server was envisioned primarily as an umbrella server (though it was also possible to post preprints directly to it). Umbrella servers always face the problem that they possess only links to papers, not the full text. So if, for example, mathematicians change jobs and move



all their papers from one departmental server to another, any umbrella server that pointed to those papers now has only dead links (and perhaps the metadata of the papers). Furthermore, the AMS preprint server was ephemeral by design: Every article posted to it had a date on which the article would be removed. The experience accumulated since the AMS preprint server was attempted has shown that reliability of access and permanence are two of the most important features of successful servers. In the end, the AMS preprint server died from lack of use. (A similar fate befell preprint servers launched around the same time by the Canadian Mathematical Society and by the American Physical Society.) Today, instead of offering its own server, the AMS maintains a comprehensive Web listing of preprint servers in mathematics (see sidebar for the URL).

Growing the Mathematics arXiv

The mathematics section of the arXiv has had its own share of growing pains. As chair of the advisory board for the mathematics arXiv, David Morrison has been deeply involved in its development. For several years he ran a preprint server in algebraic geometry that he started at Duke University in 1992 using Ginsparg's software. In the decade since then, Morrison has worked on various efforts to propagate the use of preprint servers in mathematics. One of these efforts was a collaboration between the Mathematical Sciences Research Institute (MSRI) at Berkeley and the Los Alamos server, in which preprint servers were established in each of three areas of concentration at MSRI during the 1994–95 academic year. As it turned out, only one of the three servers, that in differential geometry, took off. This server, together with the one in algebraic geometry at Duke, were eventually moved to Los Alamos. In 1997, a group of mathematicians, including Morrison, decided to form themselves into a committee to consider how to set up a mathematics section of the arXiv (this committee, with some additional members, later became the mathematics arXiv advisory board).

In the fall of 1997, the committee approached people who were running subject-based servers and asked them to move their holdings to the arXiv. About ten agreed. One was Dale Alspach of Oklahoma State University, who had started the Banach Space Archive in 1989. Alspach agreed to move his archive's approximately 340 papers to the arXiv in order to reduce his own administrative burden. He is now a moderator for the functional analysis section of the arXiv (Marc Rieffel of the University of California, Berkeley, is the moderator for the operator algebra section, and he and Alspach review postings in both sections as an additional check on the classifications). Being a moderator has allowed Alspach to stay in touch with new postings and to continue the e-mail notification service and website of the Banach Space Archive. In this way, the Banach Space Archive has become an overlay of the arXiv. "The Banach space people still exist as a group," Alspach says. "By maintaining the Web pages and the distribution list, the sense of community continues."

Others who run preprint servers decided against moving their holdings to the arXiv (though many of them encourage authors to post articles in both places). Some believe that keeping their own servers helps to raise the profile of their research areas. Some simply did not want to spend the time and effort needed to make the transition, especially when their servers were popular and functioning without any problems. But some also felt steamrollered by what they viewed as aggressive overtures of the much larger arXiv. Hans Koch, who is one of the people who run the mp_arc, says he found the arXiv committee inflexible and unwilling to compromise in their plans for merging smaller servers into the arXiv. "We would be open to the idea if there were a way to do it cooperatively and in a less centralized way…They were so sure they would be the only ones to survive, so they were not willing to compromise," he remarks. He adds, "There was a lot of pressure, and many of the smaller archives were totally run over. I think that this loss of diversity and competition is not good for our community." Morrison acknowledges that the committee inadvertently caused some resentment. "With hindsight, our self-appointed

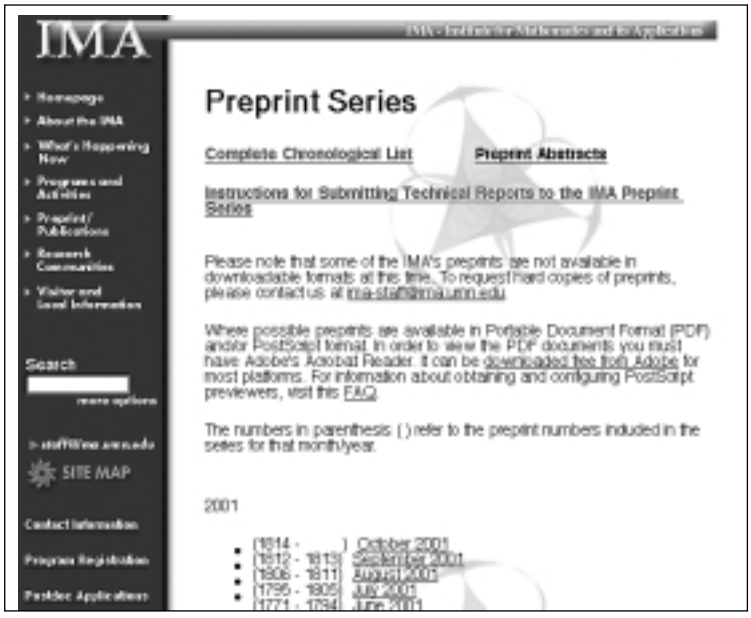
committee was not as politically astute as we could have been," he says. "We ruffled some feathers, and I'm sorry about that." He says that "the door is open" if smaller servers want to join the mathematics arXiv.

The arXiv has been successful in getting journals to agree to "direct submissions" of papers posted to the arXiv. This means an author can submit a paper to a journal by first posting the paper to the arXiv and then letting the editor know the arXiv number of the paper; in the same way, the editor can circulate the paper to referees. About forty journals now permit direct submissions. In addition, four journals are now overlay journals of the arXiv, including one of the top journals in mathematics, *Annals of Mathematics* (published jointly by the Institute for Advanced Study and Princeton University). An overlay journal not only permits direct submissions but also guarantees that all of its published papers will be posted on the arXiv. Morrison says that the advisory board has been encouraged by its interactions with journals, and the board is clearly heartened by the prestige of the *Annals* overlay. The fact that the *Annals* is using the arXiv as a repository for its papers "has influenced a number of mathematicians who had dismissed the arXiv as a place to put their papers," Morrison remarks. "They are not so dismissive anymore."

Big versus Small

"Searching for math preprints currently is very difficult," wrote Martha Tucker, a librarian at the University of Washington, in an e-mail news group. "One must try to guess which preprint server it is on, go hunt for it, not find it, guess again, etc. The current system may work fine for those working in [a] specific area, but students and non-specialists are having a difficult time. I usually end up tracking the author down instead…It would be a great idea to centralize access to all math preprints." Tucker's comments were made in 1997, but her point holds today. An umbrella server like MPRESS provides a way to find papers, but it does not have the papers themselves. At the other end of the spectrum, the centralized arXiv has a very large number of papers, but no way to get at papers not stored there. The contrast between these two kinds of servers, one distributed and the other centralized, points to a major question: Which is the best way to go?

Hans Koch believes a distributed system would be best. In much the same way that the world has moved away from centralized mainframe machines to distributed computing environments, he believes there will be a move away from huge centralized archives and toward distributed "virtual archives". A virtual archive would look to the user like one big archive, but its holdings would be distributed among smaller individual sites, such as those in mathematics departments or institutes. The individual sites could be mirrored so that if something is deleted it is not lost entirely. The sites would have to agree to be "a little satellite in this way", Koch says. He also worries that centralized archives could prove to be a drain on research funds and believes spreading the costs among individual sites is preferable. And, he says, spreading the responsibility for maintaining archival material means



that those who are closest to, and most interested in, the material are the ones who keep track of it.

Others argue that monitoring such a constellation of sites would take as much work as running a centralized archive. When it comes to archiving, says Joseph Halpern, "bigger is better". Storage costs are small compared to personnel costs, and with a big operation like the arXiv, the personnel costs are concentrated. "The same code will run whether you have 100 papers or 100,000," Halpern remarks. "So economies of scale pay off." Furthermore, small servers based in departments or institutes are sometimes at the mercy of computer administrators. Two people who run such servers and who were interviewed for this article reported that they had been hampered in doing certain upgrades on their servers because of problems with departmental computer administrators. This kind of problem does not arise with the arXiv.

Asked about the pros and cons of distributed versus centralized archives, Peter Michor says that both are needed. "We need lots of mathematical material everywhere," he remarks. "Redundancy is what we need." The arXiv "sees itself as the ultimate archive of primary mathematical material,"

he explains. By contrast, an umbrella server like MPRESS tries to coordinate material in many smaller servers. "We don't have only one big central university library on earth," he notes. "We have thousands of libraries." In the same way, the world needs many different places where it can access mathematical material.

The Three Ps: Permissions, Peer Review, and Permanence

The advent of preprint servers in mathematics has raised various concerns about this form of communication and storage of scholarly literature. Some of the concerns most often cited are: copyright permissions, lack of peer review, and long-term permanence.

The issue of copyright permissions is a murky one, for the rules vary from publisher to publisher. For example, the AMS allows all authors to retain their own copyright if they wish. Elsevier retains copyright of all material it publishes, but authors can request permission to post published journal articles on the Web; according to Elsevier's publishing director for mathematics and computer science Michiel Kolman, this permission is routinely granted.[1] Some publishers allow papers to be posted on the authors' personal websites but not on preprint servers, while others allow authors to post preprint versions of a paper, but not the final published version. Interviews for this article indicate that it is rare for a publisher to ask a preprint server to remove a published article (although sometimes authors make this request). Suppose a paper is refereed, typeset, and ready to be published, and the copyright form is sent to the author, says Richard Palais. If at that stage "the author refuses to give up the electronic rights, will the publisher say it won't publish the paper?" he asks. Most publishers would relent, he believes.

Lack of peer review of material on preprint servers has also been a concern. Because the papers have not been peer-reviewed, how can one know whether they are correct? Could it be that preprint servers are filled with erroneous papers? "This is always a concern when people first hear about the arXiv," says Michor, who is on the mathematics arXiv advisory board. "But no one using the arXiv complains about it." When an author posts a paper to the arXiv, hundreds of people immediately receive the abstract of the paper through the e-mail notification service, so authors are fairly careful about what they post. The majority of articles posted to preprint servers are submitted to journals, but no statistics exist on what percentage eventually get published. As an experiment, Greg Kuperberg looked at the publication status of the first 100 papers in theoretical high energy physics posted to the arXiv in December 1998. He found that 81 had appeared in journals, 11 were conference proceedings or invited lectures, and 2 were Ph.D. theses. "Thus at least 94 of the 100 have been blessed by some form of peer review," he concludes.

Setting aside the issue of correctness of the papers, could it be that preprint servers are filled with uninteresting junk? That 850 people subscribe to the mp_arc e-mail notification service, for example, seems to argue against the notion that this archive contains mostly uninteresting material; usage statistics for the arXiv tell the same story. One might think that preprint servers receive a lot of amateur postings with "proofs" of Fermat's Last Theorem and the like. However, those interviewed for this article say that in fact such postings are surprisingly rare; they are also easy to spot and remove. Generally, mathematicians seem to be able to navigate their way through preprint servers to locate material they find interesting. Nevertheless, most of those interviewed for this article stress the importance of maintaining a clear distinction between preprints and peer reviewed papers published in journals. Indeed, with electronic communication making greater amounts of mathematical material more widely available, journals may become even more important in imposing standards and quality control on what might otherwise be an indigestible mass of material.

The final and perhaps biggest concern about preprint servers is permanence. One of the central questions is the long-term reliability of electronic storage. Paul Ginsparg argues that this is a logistical problem, not a fundamental one. Just as librarians have been migrating material on decaying acid paper to microfilm, electronic information can be moved to new formats as the need arises. "The databases need to be actively curated, accepting only the more stable formats, and over time flagging any formats likely to become obsolete and mass migrating to newer formats," he says. "Electronic documents that are ignored for a century could certainly be problematic, and that's what has to be avoided." A compelling argument for having one big archive is that such migrations can be done far more easily and efficiently than if the data is spread around in hundreds of smaller archives.

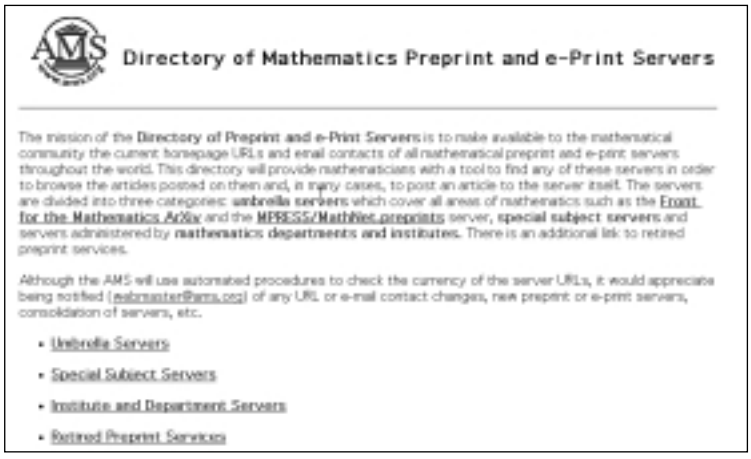
Others are less certain than Ginsparg. "I really think that this whole issue of converting to new formats has never been carefully addressed," argues Steven Krantz of Washington University in St. Louis. "I have been at meetings where people would flippantly assert that if you build 5 percent into your budget for conversions, then you are 'covered'

[1] *Elsevier publishes some journals on behalf of scientific societies, and these societies sometimes retain copyright of the journal articles. In such cases, Elsevier is not always at liberty to grant permission to post the articles. But Kolman says that when Elsevier holds the copyright, it will routinely grant this permission.*

for all eternity, which is obviously nonsense. My experience is that converting to new formats is usually a low priority, and it usually involves unforeseen complications (operating system inconsistencies, media problems, etc.)." In an opinion piece about electronic storage ("Freeware or Vaporware", *Notices*, December 2000, page 1357), Krantz mentioned several cautionary tales, such as the loss of data from the first Voyager mission in the 1970s. Although awareness of such problems has increased over the years, available solutions are often less than ideal. For example, several years ago the AMS converted its archive of published journal articles from one TEX format to another. A computer program successfully converted 90 percent of the articles, but 10 percent had to be converted by hand, requiring substantial effort by highly trained personnel. For an archive containing, say, half a million articles, that problematic 10 percent would mushroom into a huge and costly task. "I think that electronic media are very valuable for dissemination and document creation," Krantz notes. His worries center on using these media for archiving.

Another concern pertaining to permanence of electronic archives is how to ensure their financial and institutional security. This concern is sometimes raised about departmental servers, which often depend on the energy and enthusiasm of a single individual. But more often it has been raised in connection with the mathematics arXiv, which is trying to position itself to become the centralized archive for the mathematical literature. Some believe that the arXiv lacks sufficient institutional and financial support to guarantee its long-term existence (though the move from Los Alamos to Cornell University should provide greater stability). In response to this concern, one suggestion made is that a consortium of scientific professional societies could band together to provide long-term support for the arXiv. John Ewing, executive director of the AMS, believes this idea is unrealistic at the moment because the long-term costs of supporting a large electronic archive are unknown. Also, sharing the editorial, administrative, and technical decisions among a number of societies could prove to be an organizational nightmare. Ewing also worries about a single-minded focus on the arXiv at a time when scholarly communication is in so much flux. "No one knows what works best now, so it's extremely important to keep lots of different things going at the same time, until we know better what will develop," he says.

What is the future for preprint servers in mathematics? Many are convinced that in a decade or so, nearly all of the current mathematical literature will be accessible this way. Electronic archives may also eventually contain a portion of the peer-reviewed literature, especially that from electronic



journals. Some dream of the day when MathSciNet will contain links not only to electronic versions of peer-reviewed articles but to articles on preprint servers as well. And some are also convinced that a new breed of search engines, ones that can understand mathematical content, will revolutionize the electronic navigation of mathematical literature.

Over the next several years, the mathematical community will have to come to grips with many of the issues raised by the increasing use of preprint servers. There is no road map that can provide directions toward the most successful route, and the paths broken by other sciences may not be the best ones for mathematics. Thus mathematicians themselves must guide the evolution of electronic archives. As Laurent Siebenmann, a mathematician at the Université de Paris-Sud, put it in an e-mail discussion group:[2] "We have a baby in our care—the electronic math literature—and the vital task is to mother it *today*."

[2] *Electronic Math Journals Discussion List,* EMJ@listserv.albany.edu, *archived at* http://math.albany.edu:8800/hm/emj/.