# First Steps Towards Electronic Research Communication

Paul Ginsparg

Citation: Computers in Physics **8**, 390 (1994); doi: 10.1063/1.4823313
View online: https://doi.org/10.1063/1.4823313
View Table of Contents: https://aip.scitation.org/toc/cip/8/4
Published by the American Institute of Physics

# First Steps Towards Electronic Research Communication

**Paul Ginsparg**

***Barely three years old, the e-print archives at Los Alamos National Laboratory now serve a user base of more than 20,000***

The e-mail address `hep-th@xxx.lanl.gov`, for the first of a series of automated archives for electronic communication of research information, went online starting in August 1991. This "e-print archive" began as an experimental means of circumventing recognized inadequacies of research journals but unexpectedly became within a very short period the primary means of communicating ongoing research information in formal areas of high-energy particle theory. Its rapid acceptance within this community depended critically both on recent technological advances and on behavioral aspects of this research community, as described in this article. There are now more than 3600 regular users worldwide of `hep-th`—which stands for high-energy physics theory. The archiving software has also been expanded to serve a total of 14 different research disciplines:

- High-energy particle theory (formal): started August 1991
- Algebraic geometry: started February 1992
- High-energy particle theory (phenomenological): started March 1992
- Astrophysics: started April 1992
- Condensed-matter theory: started April 1992
- Computational and lattice physics: started April 1992
- Functional analysis: started April 1992
- General relativity/Quantum cosmology: started July 1992
- Nuclear theory: started October 1992
- Nonlinear sciences: started March 1993
- Economics: started July 1993
- High-energy experimental physics: started April 1994
- Chemical physics: started April 1994
- Computation and language: started April 1994

Combined archive usage has grown dramatically (see Fig. 1).

The extended, automatically maintained database and distribution system serves over 20,000 users from more than 60 countries and processes over 30,000 messages per day; it is already one of the largest and most active databases on the Internet. This system provides a paradigm for recent changes in worldwide, discipline-wide scientific information exchange and a model for electronic transmission of research and other information when the next generation of "electronic-data highways" begins to provide more universal access to high-speed computer networks.

## Background

In my own community of high-energy theoretical physics, the rapid acceptance of electronic communication of research information was facilitated by a pre-existing "preprint culture," in which the irrelevance of refereed journals to ongoing research has long been recognized. At least since the mid-1970s, the primary means of communicating new research ideas and results had been a preprint-distribution system in which printed copies of papers were sent via ordinary mail to large distribution lists at the same time that they were submitted to journals for publication. (Larger high-energy-physics groups typically spent between \$15,000 and \$20,000 per year on photocopy, postage, and labor costs for their preprint distribution.) These papers could then typically take six months to a year to be published and appear in a journal. In this community, we have therefore learned to determine from the title and abstract (and occasionally the authors) whether we wish to read a paper, and to verify necessary results rather than rely on the alleged verification of overworked or otherwise careless referees. The small amount of filtering provided by refereed journals plays no effective role in our research.

In addition, this community, by the mid-1980s, had already begun highly informal mechanisms of regular electronic information exchange, in turn enabled by concurrent advances in computer software and hardware. The first such advance was the widespread standardization during this period on TeX (written by Donald E. Knuth of Stanford) as our scientific word processor. For the first time, we could produce for ourselves a printed version equal or superior in quality to the ultimate published version. TeX has in addition the virtue of being ASCII-based, which makes transmitting TeX files between different computer systems straightforward. Collaboration at a distance became extraordinarily efficient, because we no longer had to express-mail multiple versions of a paper back and forth but could instead see one another's

*Paul Ginsparg is a theoretical physicist and staff member at Los Alamos National Laboratory, NM 87545. E-mail: ginsparg@lanl.gov*

revisions essentially in real time. Figures can also be generated within a TeX-oriented picture environment or, more generally, can be transmitted as standardized Postscript files produced by a variety of graphics programs.

A second technological advance was the exponential increase in computer network connectivity achieved during the same period. By the end of the 1980s, virtually all researchers in this community were plugged into one or another of the interconnected worldwide networks and were using e-mail on a daily basis. (Conceivably, earlier attempts to set up "electronic journals" in other communities failed not due to any intrinsic flaw in implementation, but rather due to the insufficiently mature state of computer networking itself—leaving an underconnected userbase.) Finally, the existence of online archives that allow access to large amounts of information has been enabled by the widespread availability of low-cost but high-powered workstations with high-capacity storage media. An average paper (with figures, and after compression) requires 40 kbytes to store. Hence, one of the current generation of rapid-access gigabyte disk drives costing under $1000 can hold 25,000 papers at an average cost of $0.04 per paper. Slower-access media for archival storage cost even less: a digital audio tape cartridge, available from discount electronics dealers for under $15, can hold over 4 Gbytes, that is, over 100,000 such papers. The data equivalents of multiple years of most journals constitute a small fraction of what many experimentalists routinely handle on a daily basis, and the costs of data storage will only continue to diminish.

*This system provides a paradigm for recent changes in worldwide, discipline-wide scientific information exchange.*

Because storage is so inexpensive, it can be duplicated at several distribution points, minimizing the risk of loss from accident or catastrophe and facilitating worldwide network access. The Internet runs 24 hours per day with virtually no interruptions and transfers data at rates up to 45 Mbits/s (that is, less than 0.01 s per paper). Currently projected upgrades of NSFnet to a few gigabits per second within a few years should be adequate to accommodate increased usage for the academic community. Commercial networks currently envisioned for the nation's electronic data highway will have even greater capacity.

The above technological advances—combined with a remarkable lack of initiative on the part of conventional journals in response to the electronic revolution—rendered the development of e-print archives "an accident waiting to happen." Perhaps more surprising has been the readiness of scientific communities to adopt this new tool of information exchange and to explore its implications for traditional review and publication processes. The exponential growth in archive usage suggests that scientific researchers are not only eager, but indeed impatient, for completion of the proposed "information super-highways" (though not necessarily the "information turnpikes").

## Implementation

Having concluded that an electronic preprint archive was possible in principle, I spent a few afternoons in the summer of 1991 writing the original software to assess the feasibility of fully automating such a system. At issue was whether an automated information exchange could permit its users to construct, maintain, and revise a comprehensive database and

`xxx.lanl.gov`, *currently an HP 9000/735, delivers "e-prints" day and night from the floor under a table in a corner of the author's office. This modest installation responds automatically to 30,000 messages per day, serving a user base consisting at present of more than 20,000 scientists in 60 countries.*

distribution network without outside supervision or intervention. The software was rudimentary and allowed users with minimal computer literacy to communicate e-mail requests to the Internet address **hep-th@xxx.lanl.gov**. Remote users could submit and replace papers, obtain papers and listings, get help on available commands, search the listings for author names, and so on. The system allows ongoing corrections and addenda and is implemented to assure that only those who so desire are kept up to date.

It is important to distinguish the formal communication of such an "e-print archive" (which meets the standards and needs of the scientific community for research promulgation and circulation) from the informal (and unarchived) communication provided by electronic bulletin boards and network news. In the former case, researchers are deliberately restricted to communication via their abstracts and research papers, which are in principle equally suitable for publication in conventional research journals, whereas the latter case is more akin to ordinary conversation or written correspondence, in that the communication is neither indexed for retrieval nor stored indefinitely. The e-print archives, designed as a tool for the electronic communication of research results, include features such as the aforementioned ability of a submitter to replace his or her submission; checks on database integrity (to ensure, for example, that the individual replacing a submission is indeed the original submitter); permanent records of submissions together with dates submitted; and records of number of user requests for each paper.

Users can subscribe to the system and thereby receive a daily listing of titles and abstracts of new papers received (see Fig. 2). This primitive e-mail interface provides a necessary lowest common denominator that ensures the widest possible access to users, irrespective of their network connection and local operating system. The initial user base for **hep-th** was assembled from pre-existing e-mail distribution lists in the subject of two-dimensional gravity and conformal field theory. Starting from a subscriber list of 160 addresses in mid-August 1991, **hep-th** quickly grew within six months to encompass most of formal quantum field theory and string theory; as mentioned above, it currently has over 3600 subscribers. Its smooth operation has transformed it from its initial incarnation as a feasibility-assessment experiment for a small subcommunity into an essential research tool for its users, many of whom have reported their dependence on receiving multiple "fixes" per day. The original **hep-th** archive alone now receives roughly 200 new submissions per month, responds to more than 700 e-mail requests per day, and transmits more than 1000 copies of papers on peak days. Internet e-mail access time is typically a few seconds. The system originally ran as a background job on a small Unix workstation (a 25-MHz NeXTstation with a 68040 processor purchased for roughly $5000 in 1991), which was primarily used for other purposes by another member of my research group, and placed no noticeable drain on CPU resources. The system has since been moved to an HP 9000/735 that sits exiled on the floor under a table in a corner (see photo on p. 391).

For those directly on the Internet, the system allows anonymous file-transfer protocol (FTP) access to the papers and the listings directories. Recently WorldWideWeb access has also been set up to provide an even more sophisticated and convenient mode of network access for those with the required (public-domain) client software (see Fig. 3). This

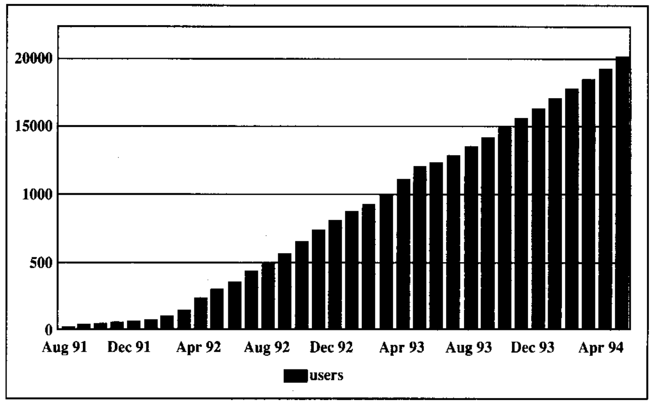


*Figure 1. Number of e-print users continues to increase rapidly. Bar chart shows combined number of users for the 14 e-print archives listed in the text.*

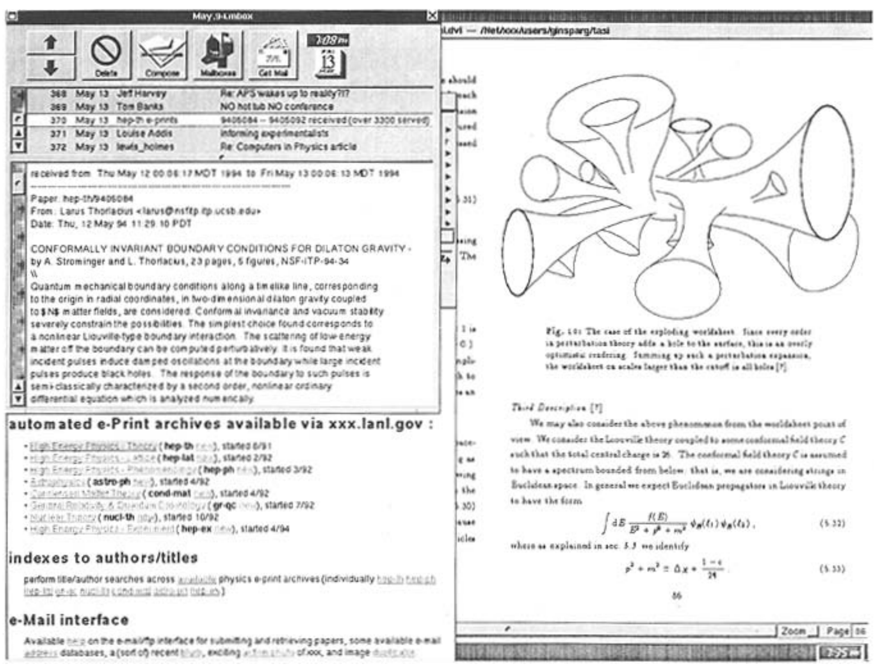


*Figure 2. Users can access the e-print archives via e-mail or through more sophisticated communication interfaces. The window in the upper left corner shows abstracts received via e-mail, and the window in the lower left corner shows the graphical user interface provided by a WorldWideWeb client (in this case* **OmniWeb.app** *running under NeXTstep) accessing the frontpage* **http://xxx.lanl.gov/** *(the underlined text signifies network hyperlinks which bring up new hypertext when clicked upon). A paper extracted from the e-print archive appears in the window on the right side, where it can be read or sent to a printer.*

software provides local menu-driven interfaces, automatically connected to the nearest central server, which transparently pipe selected papers through text formatters directly to a screen previewer or printer. Moreover, on many local networks, such software has been set up to cache and redistribute papers. The WorldWideWeb interface alone currently processes over 5000 requests daily to the multiple archives served by **xxx.lanl. gov** (see Fig. 4). Although WorldWideWeb still represents only a small fraction of the overall usage, this access mode is expected to become dominant in the near future.

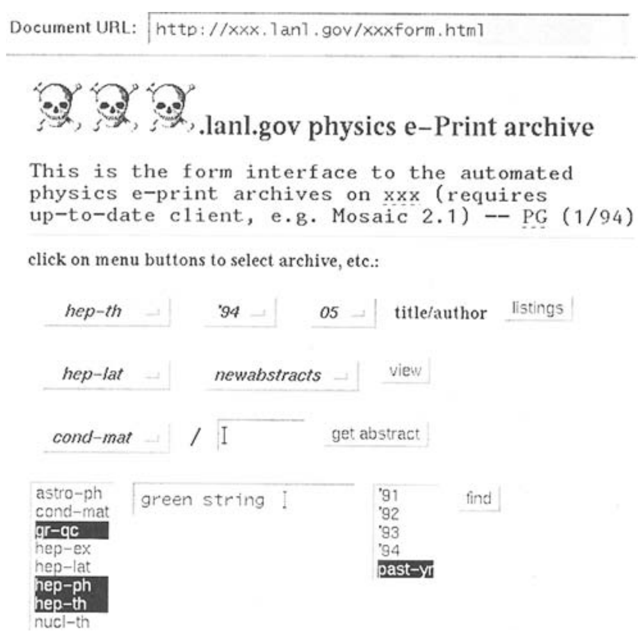


*Figure 3. Another WorldWideWeb client, Mosaic, summons this "form interface" on* **xxx.lanl.gov**. *The motif "buttons" allow the user to choose an archive to view monthly listings or daily abstracts received, or to search the title/author listings of selected archives for given time periods. Listings are displayed in hypertext with included hyperlinks that retrieve paper abstracts or full text in either TeX or Postscript format.*

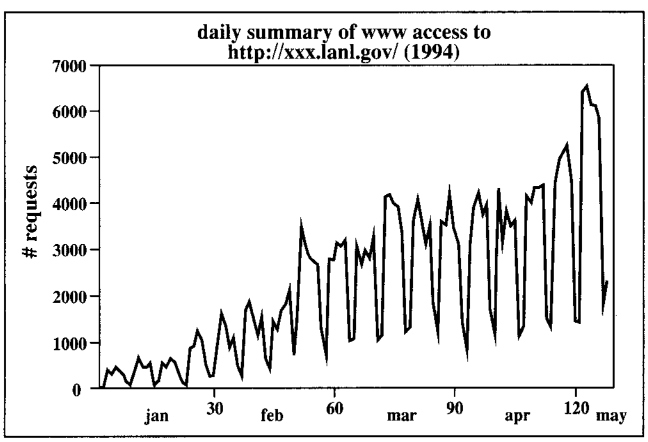


*Figure 4. The number of http (hypertext transfer protocol) requests to the WorldWideWeb interface on* **xxx.lanl.gov** *per day starting from Jan. 1, 1994, has grown to over 5000 daily. The seven-day periodicity is evidence that many physicists still do not have readily available network access on weekends.*

An active archive such as **hep-th** requires about 70 Mbytes per year (that is, \$70 per year) for storing papers, including figures. Its network usage is less than $10^{-4}$ of the **lanl.gov** backbone capacity, and so it places a negligible drain on local network resources. It requires little intervention and has run entirely unattended for extended periods while I have been away on travel. It is difficult to estimate the potential for dedicated systems of the future only because the resources of the current experimental one (run free of charge) are so far from saturation. In the meantime, additional e-print archives have been established for other disciplines of physics and mathematics (see Fig. 1), including high-energy phenomenology, general relativity, nuclear theory, algebraic geometry, condensed-matter theory, lattice physics, astrophysics, nonlinear sciences, economics, and, most recently, chemical physics and computational linguistics.

Storage and retrieval of figures have not been a problem. While software for figures has not yet been standardized, the vast majority of networked physics institutions have screen previewers and laser printers that display and print Postscript files created by a wide variety of graphics programs. Figure files are typically submitted as compressed Postscript, and papers can be printed with the figures embedded directly in the text. High-resolution digital scanners will soon become as commonplace as fax machines and permit inclusion of figures of arbitrary origin. (It is already possible to fax arbitrary figures to a machine equipped with a fax modem, convert them to bitmapped Postscript, and postpend them.) With appropriate data compression and Postscript conversion, figures typically increase paper-storage requirements by an inconsequential factor of two.

Some measure of the success of e-print archives is given, first, by widespread testaments from users that they find it an indispensable research tool—effectively eliminating their reliance on conventional print journals (and some report no longer submitting to journals, either because they have unconsciously forgotten to do so, since the information has already been communicated, or because they have consciously chosen not to deal with a tedious and seemingly unnecessary process); second, by decisions of numerous institutions to discontinue their preprint mailings in recognition of the superior service provided by e-print archives; and third, by the practice that has now become customary in some of the fields served by e-print archives of providing as reference a paper's e-print index number rather than a local report number or a published reference.

## Prospects and concerns

The system in its present form was not intended to replace journals but only to organize a haphazard and unequal distribution of electronic preprints. It is increasingly used as an electronic journal, however, because it is evidently more convenient to retrieve electronically a computer file than to retrieve physically a paper from a file cabinet. Besides minimizing geographic inequalities by eliminating the boat-mail gap between continents, the system institutes a form of democracy in research wherein access to new results is granted equally to beginning graduate students and seasoned operators. No longer is it crucial to have the correct connections or to be on exclusive mailing lists to be kept informed of progress in one's field. The pernicious problem of lost or stolen preprints experienced by some large institutions is definitively exorcised as well. Communication among colleagues at the same institution may even be enhanced, because they frequently cross-request one another's preprints from the remote server (the reasons for which in general I hesitate to contemplate). Many institutions have already eliminated their hardcopy distribution of preprints and thus have already seen significant savings in time and money; others have begun to request specifically that hard copy no longer be sent to them, since the electronic distribution has proven reliable and more efficient.

It is straightforward to implement charges for such a system if desired, via either flat access rates or monitored usage rates. Piggybacked on existing network resources, however, such systems cost so little to set up and maintain that they can be offered virtually free. Overburdened terminal resources at libraries are not an issue, because access is typically via the terminal or workstation on one's desk or in the nearest computer room.

These systems allow users to insert interdisciplinary pointers to their papers stored in related archives, thus fostering interdisciplinary distribution and retrieval of information through cross-linked databases. Electronic research archives will prove especially useful for new and emerging interdisciplinary areas of research for which there are no pre-existing journals and for which information is consequently difficult to obtain. In many such cases, it is advantageous to avoid a proliferation of premature or ill-considered new journals. Cross-linking provides an immediate virtual meeting ground for researchers who would not ordinarily communicate with one another and who can quickly establish their own dedicated electronic archive and ultimately disband if things do not pan out, all with infinitely greater ease and flexibility than is provided by current publication media.

In the long term, electronic access to scientific research will be a major boon to developing countries, since the expense of connecting to an existing network is infinitesimal compared with that of constructing, stocking, and maintaining libraries. (Indeed I frequently receive messages from physicists in developing countries confirming how much better off they find themselves even in the short term with the advent of the current electronic distribution systems: no longer are they "out of the loop" for receipt of information. Others report the feeling that their own research gets a more equitable reading and is no longer dismissed for superficial reasons of low-quality print or paper stock.) The trend experienced over the past decade in the Western World, where data-transmission lines have become as common as telephone service, and terminals and laser printers have become as common as typewriters and photocopy machines, could be repeated even more quickly as countries in eastern Europe and the Third World develop electronic infrastructures. In the short term, these countries will certainly not be worse off than they already are, receiving information via conventional means from the nearest redistribution point. Conformity to a uniform computer standard both in the United States and abroad to communicate results to the largest possible audience should pose no greater a burden than communication using a non-native language—English—already imposes on the majority of the world. Similar comments apply equally to less-well-endowed institutions in the United States. Moreover, the infrastructure trends experienced by physics and biology departments are soon to be repeated by the full range of conventional academic institutions, including teaching hospitals, law schools, and humanities departments, and ultimately as well by public libraries and K-12.

Publication companies, on the other hand, have been somewhat irresponsible over the past decade, increasing the number of journals as well as the subscription price per journal (some single journal subscriptions to libraries now cost well over $10,000 per year) during a period when libraries have experienced continuously decreasing resources and space. In the meantime, these same publication companies have been slow to incorporate electronic communication into their operation and distribution, which could ultimately result in dramatic savings in cost and efficiency for all involved. There remain numerous "value-added" enhancements (discussed below) that could vastly improve upon what is possible in the current automated archives. Perhaps these could still be provided by conventional publication companies if they are prepared to move quickly enough.

Concerns about interference from malevolent hackers have proven unfounded to date. Archives can be rendered highly resistant to corruption, and minimal precautions can assure users of remote database systems that their own system resources are not endangered. Anonymous FTP servers running on the Internet have for years allowed the academic community an open exchange of executable software far more susceptible to malfeasance, and their safeguards have proven effective. At this writing, there has happily not been a single instance of attempted break-in or deliberate destruction. (Perhaps the recent rapid growth of the Internet has simply provided too many easy targets, and random electronic vandalism has lost any perceived attraction.)

Some members of the community have voiced their concern that electronic distribution will somehow increase the number of preprints produced, or encourage dissemination of preliminary or incorrect material, arguing that an electronic version is somehow more ephemeral than a printed version and therefore more readily distributed. This concern, however, confuses the method of production with the method of distribution, and it is likely that most researchers are already producing at saturation. Secondly, posting a preprint to an electronic archive instantly publicizes it to thousands of people, and so the embarrassment over incorrect results and the consequent barrier to distributing material prematurely are, if anything, increased. Such submissions cannot be removed but can only be replaced by a note that the work has been withdrawn as incorrect, leaving a more permanent blemish than a hard copy of limited distribution, which is soon forgotten.

This is not to argue that refereed or moderated forums are obsolete or necessarily undesirable. In some disciplines, the refereeing process is regarded as playing a useful role both in improving the quality of published work and in filtering out large amounts of irrelevant or incorrect material for the benefit of readers. The refereeing process plays the additional role of validating research for the purpose of job and grant allocation. (It is useful to observe, however, that if it is citations alone that are used as a criterion for influence of research, then these can as usefully be compiled in an unrefereed as in a refereed sector: "crackpot" papers neither make nor receive enough systematic references in the mainstream literature to alter any "signal" in this methodology.) A refereeing mechanism could be easily implemented for the e-print archives in the form of either a filter prior to electronic distribution or a review ex post facto by volunteer readers and selected reviewers. In either case, the archives could be partitioned into one or more levels of refereed as well as unrefereed sectors. Thus lifting the artificial financial constraints to dissemination of information and decoupling it from the traditional refereeing process will allow for more innovative methods of identifying and validating significant research.

Additional problems may arise as computer networking spreads outside the academic community. For example, **hep-th** would be somewhat less useful if it were to become inundated for example by submissions from "crackpots" promoting their perpetual-motion machines or even from well-meaning high-school students claiming to refute the special theory of relativity, and so on. Perhaps such submissions will ultimately constitute no greater a nuisance than is currently experienced by recognized journals, or become no more commonplace an annoyance than currently are unwanted physical or telephone intrusions into our offices and homes. It is clear, however, that the architecture of the data highways of the future will somehow have to reimplement the protective physical and social isolation now enjoyed by ivory towers and research laboratories.

Increased standardization of networking software and electronic storage formats during the 1990s encourages us to fantasize about other possible enhancements to scholarly research communication. Usenet newsgroups, for reasons such as their lack of indexing and archiving, and excessively open nature, are unlikely to prove adequate for serious purposes. On the other hand, it is now technically simple to implement for example a WorldWideWeb form-based submission system to build hyperlinked discussion threads, accessible linked from individual papers and also from a subject-based discussion page. All posted text could be WAIS-indexed for easy retrieval, and related threads could interleave and cross-link in a natural manner, with standard methods for moving forward and backtracking. A histogram-like interface would facilitate finding active threads, and the index could allow location of all postings by a given person (including oneself) with date of latest follow-up to facilitate tracking of responses.

A hyperlinked system would provide much more flexibility than Usenet, specifically avoiding awkward protocols for group creation and removal and also avoiding potentially unscalable aspects of nntp (the network news transfer protocol). For the relatively circumscribed physics-research community, a central database (with the usual mirrored nodes) would have no difficulty with storage or access bandwidth. To enable full-fledged research communication with in-line equations or other linkages, we require slightly higher-quality browsers than are currently available. But with hypertext transfer protocols (http) now relatively standardized, network links and links to other application software can be built into underlying TeX (and configured into standard macro packages) to be interpreted either by dedicated TeX previewers or passed by a suitable driver into more archival formats (such as Adobe Acrobat pdf) for greater portability across platforms. Multicomponent messages could also be assembled in a graphical MIME (multipurpose Internet mail extension) composing object to be piped to the server via the http POST protocol, thereby circumventing some of the inconvenient baggage of Internet sendmail or FTP protocols.

While the above is technically straightforward to implement, there remains the aforementioned issue of limiting access to emulate that effective insulation from unwanted incursions afforded by corridors and seminar rooms at universities and research laboratories. One method would be to employ a "seed" mechanism, that is, to start from a given set of "trusted" users and let them authorize others (and effectively be responsible for those beneath them in the tree), with guidelines for admission based on educational level or other criteria. Permission to post or authorize would be revokable at any time, retroactive one level back in the tree. To allow global coverage, application to the top level for authorization could be allowed to start a new branch. The scheme entails some obvious compromises, and other schemes are easily envisioned, but the ultimate object remains to determine the optimal level of filtering for input access to maintain an auspicious signal/noise ratio for those research communities that prefer to be buffered from the outside world. This would constitute an incipient "virtual communication corridor," further facilitating useful research communication in what formerly constituted both pre- and post-publication phases, and rendering ever more irrelevant individual researchers' physical location.

Finally, we mention that the e-print archives in their current incarnation already serve as surprisingly effective inducements in the campaign for computer literacy and have succeeded in motivating some dramatic changes in computer usage. Researchers who previously disdained computers now confess an addiction to e-mail; many who for years had refused to switch to Unix or to TeX are in the process of converting; others have suddenly discovered the power of browsing with WorldWideWeb. The effectiveness of the systems in motivating these changes justifies the philosophy of providing dual functionality in the form of top-of-the-line search, retrieval, and input capabilities for cutting-edge "power" users, while maintaining "lowest common denominator" capabilities for the less network-fortunate.

## Conclusions and open questions

These systems are still primitive and are only tentative first steps in the optimal direction. To summarize, thus far we have learned:

- The exponential increase in electronic-networking usage over the past few years opens new possibilities for both formal and informal communication of research information.
- In some fields of science, electronic preprint archives have been on-line since mid-1991 and have become the primary means of communicating research information to many thousands of researchers within the fields they serve. It has been established that people will voluntarily subscribe to receive information from these systems and make aggressive use of them if they are set up properly. Such systems are expected to grow and evolve rapidly in the next few years.
- From such experimental systems, we have learned that open (that is, unrefereed) distribution of research information can work well for some disciplines and has advantages

for researchers both in developed and developing countries. We have also learned that the technology and network connectivity are currently adequate to support such systems, whose performance should benefit from continuous technological improvement expected in the near future.

I conclude with some unanswered questions to amplify some of my earlier comments:

- Who will ultimately be the prime beneficiaries—researchers, publishers, libraries, or other network-resource providers—of electronic research communication?
- What factors influence research communities in their rate and degree of acceptance of electronic technology, and what mechanisms are effective in facilitating such changes?
- What role will be played by the conventional peer-refereeing process in the electronic media, and how will it differ from field to field?
- What role will be played by publishing companies, and how large will their profits be? If publication companies do adopt fully electronic distribution, will they pass along the reduced costs associated with the increased efficiency of production and distribution to their subscribers? Can publishing companies provide more value-added than an unmanned automated system, the primary virtue of which is instant retransmission?
- What role will library systems play? Will information be channelled somehow through libraries, or directly to researchers?
- How will copyright law be applied to material that exists only in electronic form? At the moment publishing companies have "looked the other way" from these systems, living with the dissemination of the electronic preprint information as they did with the earlier preprinted form and claiming that it would be antithetical to their philosophy to impede dissemination of information. Will they continue to be so magnanimous when libraries begin to cancel journal subscriptions?
- What storage formats and network utilities are best suited for archiving and retrieving information? Currently we use a combination of e-mail, anonymous FTP, and window-oriented utilities such as Gopher and WorldWideWeb combined with WAIS indexing to retrieve TeX and Postscript documents. Will something even better, such as Acrobat or some other format currently under development, soon merge with the above or emerge as a new standard?
- How will the medium itself evolve? Conservatively we can imagine "interactive" journals in which equations can be manipulated, solved, or graphed; in which citations can instantly open references to the relevant page; and in which comments and errata dated and keyed to the relevant text can be inserted as electronic "post-it" notes in the margins. Ultimately we will have a multiply interconnected network hypertext system with transparent pointers among distributed databases that transcends the limits of conventional journals in structure, content, and functionality, thereby transforming the very nature of what is communicated. This is the sort of "value added" for which we should certainly be willing to pay. Certainly we do not wish to clone current journal formats (determined as they are by constraints particular to the print medium) in the electronic medium; we are already capable of distinguishing information content from superficial appearance. Who will decide the standards required to implement any such progress?

This began for me as a spare-time project to test the design and implementation of an electronic preprint distribution system for my own relatively small research community. Its feasibility had been the subject of contentious dispute, and its realization was thought even by its proponents to be several years in the future. Its success has led to an unexpectedly enormous growth in usage; it has expanded into other fields of research and has in addition elicited interest from many others. (I have received over a hundred inquiries into setting up archives for different disciplines.) Each discipline will have slightly different hardware and software requirements, but the current system can be used as a provisionary platform tailorable to the specific needs of different communities. Nonetheless it has remained a spare-time project with little financial or logistical support.

Further development will require coordination among interested researchers from various disciplines, computer and networking staff, and interested library personnel, and in particular will require dedicated staffing. At the moment, hardware and software maintenance of existing automated archives remains a loosely coordinated volunteer operation, and little further progress can be made on the issues raised by the current systems without some thoughtful direction. Perhaps the centralized databases and further software development will ultimately be administered and systematized by established publishing institutions, if they are prescient enough to reconfigure themselves for the inevitable. Since it has been researchers who have taken the lead thus far, however, we should retain this unique opportunity to continue to lead the development of such systems in optimal directions and on terms maximally favorable to ourselves.

## Acknowledgments

Many people have contributed (consciously or otherwise) to the development of these systems. The original distribution list from which **hep-th** sprang in 1991 was assembled by Joanne Cohn, whose incipient efforts demonstrated that members of the high-energy-physics community were anxious for electronic distribution (and Stephen Shenker recommended that the original archive name not include the string "string"). Continual improvements have been based on user feedback too voluminous to credit (although among the most vocal have been Tanmoy Bhattacharya, Jacques Distler, Marek Karliner, and Paul Mende). People who have administered some of the remotely based archives include Dave Morrison, Bob Edwards, Roberto Innocente, Erica Jen, and Bob Parks. Joe Carlson and David Thomas set up the original gopher interfaces in late 1992. The Network Operations Center at Los Alamos National Laboratory has reliably and uncomplainingly supplied the requisite network bandwidth **@lanl.gov**, and Joe Kleczka has been available for crisis control. Louise Addis and staff at the SLAC library moved quickly to incorporate e-print information into the SPIRES database, furthering their decades of tireless electronic service to the high energy physics community. Finally, Geoffrey West repeatedly and against all obvious reason insisted that it is appropriate for Los Alamos National Laboratory to sponsor this activity, while simultaneously bearing the bad news both from within the laboratory and from certain government funding agencies.