P. Ginsparg 19 Apr 1995

First Steps towards Electronic Research Communication *
(http://xxx.lanl.gov/)

hep-th@xxx.lanl.gov, the e-mail address for the first of a series of automated archives for electronic communication of research information, went on-line 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 below. There are now more than 3800 regular users of hep-th worldwide, and the archiving software has also been expanded to serve over 25 other research disciplines, including not only other areas of high energy physics such as phenomenology, lattice computations, and experiment; but also to other areas of physics such as astrophysics, condensed matter theory, materials theory, general relativity & quantum cosmology, nuclear theory & experiment, quantum physics, chemical physics, accelerator physics, superconductivity, and plasma physics; and as well to fields such as non-linear dynamics and Bayesian analysis, certain fields of mathematics including algebraic geometry, functional analysis, quantum algebra, differential geometry, automorphic forms, and complex dynamics; and even to disciplines such as economics, computation & linguistics, and atmospheric & oceanic sciences (WorldWideWeb access to these archives is available at http://xxx.lanl.gov/). The extended automatically maintained database and distribution system serves over 25,000 users from more than 70 countries, and processes over 45,000 electronic transactions 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:

The rapid acceptance of electronic communication of research information in my own community of high energy theoretical physics 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 communication of 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 \$15k and \$20k per year on photocopy, postage, and labor costs for their

* Adapted from "Computers in Physics" volume 8, no. 4, Jul/Aug 1994, pp. 390-396.

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.

This community, by the mid 1980's, had as well 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 wordprocessor. 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 it straightforward to transmit TeX files between different computer systems. Collaboration at a distance became extraordinarily efficient, since we no longer had to express-mail multiple versions of a paper back and forth and could instead see one another's revisions essentially in real time. Figures as well can 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 1980's, 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. (It is conceivable that 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 which 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 kilobytes to store. Hence one of the current generation of rapid-access gigabyte disk drives costing under $600 can hold 25 000 papers at an average cost of less than 3 cents 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 gigabytes, 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.

Since the storage is so inexpensive, it can be duplicated at several distribution points, minimizing the risk of loss due to accident or catastrophe, and facilitating worldwide network access. The internet runs 24 hours a day with virtually no interruptions, and transfers data at rates up to 45 megabit/sec (i.e., less than .01 sec per paper). Currently projected

backbone upgrades 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 to comprise 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 superhighways" (though not necessarily the "information turnpikes").

IMPLEMENTATION:

Having concluded that an electronic preprint archive was possible in principle, in the summer of 1991 I spent a few afternoons 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 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 abovementioned Internet address hep-th@xxx.lanl.gov ("hep-th" stands for high-energy physics theory). 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, i.e. 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/her submission; checks on data base 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. This primitive e-mail interface provides a necessary lowest

common denominator that insures the widest possible access to users, independent 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, and as mentioned above currently has over 3800 subscribers (still the largest individual archive, though a small percentage of the over 25000 subscribers aggregated over all of the archives). 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, and the archives as a whole process roughly 350 new submissions processed per week. 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 \$5k 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. (A photo of the setup and related figures illustrating the growth in usage together with current access rates are available via WorldWideWeb at http://xxx.lanl.gov/blurb/)

The system also allows anonymous FTP access to the papers and the listings directories, and WorldWideWeb access provides an even more sophisticated and convenient network access for those with the required (public-domain) client software. This allows local menu-driven interfaces, automatically connected to the nearest central server, that transparently pipe selected papers through text formatters directly to a screen previewer or printer. (Such software has moreover been set up to cache and redistribute papers on many local networks.) The WorldWideWeb interface alone on xxx.lanl.gov currently processes roughly 15000 requests per day on weekdays, and has been growing each month by an additional 30,000 requests (as of Apr 95).

An active archive such as hep-th requires about 70 megabytes per year (that is, \$40 per year) for storing papers, including figures. Its network usage is less than 10^{-4} of the lanl.gov backbone capacity, so places a negligible drain on local network resources. In principle 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 the more than 25 other disciplines of physics, mathematics, and more distant displines mentioned at the outset here.

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 of course possible to fax arbitrary figures to a machine equipped with a fax modem, convert to bitmapped Postscript, and postpend them.) With appropriate data compression and Postscript conversion, figures typically increase paper storage requirements by an inconsequential factor of 2.

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, since the information is already communicated, or because they have consciously chosen not to deal with a tedious and seemingly unnecessary process); second, decisions by numerous institutions to discontinue their preprint mailings in recognition of the superior service provided by e-print archives; and third, that it has now become customary in some of the fields served by e-print archives to provide as reference a paper's e-print archives 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 of a paper 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 as well definitively exorcised. Communication among colleagues at the same institution may even be enhanced, since 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, since 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 wouldn't ordinarily communicate with one another, who can then 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, 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 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 they are certainly no 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 US 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 those less well-endowed institutions in the U.S., and the trends experienced by physics and biology departments are soon to be repeated by the full range of conventional academic institutions, including e.g. 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 and as well the subscription price per journal (some single journal subscriptions to libraries now run well over $10k 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 (some discussed further below) that could vastly improve

upon what is possible in the current automated archives. The resources necessary for production and distribution of conventional printed journals formerly allowed publishers to focus on only those mechanics, avoiding any pressure to rethink the intellectual content and quality of their operations. But the on-line electronic format will allow us to transcend the current inadequate system for "validating" research in a variety of ways. No longer need we be tied to a one-time all-or-nothing referee system which provides insufficient intellectual signal, and a static past database. We eagerly anticipate a vastly improved and more useful electronic literature, taking advantage of the flexibility afforded by the electronic medium and unhindered by artifacts of its evolution from paper.

Moreover, it is difficult to imagine how the current model of funding publishing companies through research libraries (in turn funded by overhead on research grants) can possibly persist in the long term. It is premised on a paper medium that was difficult to produce, difficult to distribute, difficult to archive, and difficult to duplicate – a medium that hence required numerous local redistribution points in the form of research libraries. The electronic medium shares none of these features and thus naturally facilitates largescale disintermediation, with attendant increases in efficiency benefitting both researchers and their sources of funding. Recent developments have exposed the extent to which current publishers have defined themselves in terms of production and distribution, roles which we now regard as trivially automated. But there remains a pressing need for organization of *intellectual* value-added, which by definition cannot be automated even in principle, and that leaves significant opportunities for any agency willing to listen to what researchers want and need.

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 databases systems that their own system resources are not endangered. Anonymous FTP servers running on 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 malfeasance. (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 lowers their barrier to distributing. 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, the electronic form once posted to an archive is instantly publicized to thousands of people so the embarrassment over incorrect results and consequent barrier to distributing is, 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 to play 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/or 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 of the academic community. hep-th for example would be somewhat less useful if it were to become inundated for example by submissions from "crackpots" promoting their perpetual motion machines, or even well-meaning high school students claiming to refute the special theory of relativity, and so on. Perhaps this 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 information data highways of the future will somehow have to reimplement the protective physical and social isolation currently enjoyed by ivory towers and research laboratories.

Increased standardization of networking software and electronic storage formats during the 1990's 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 given points in individual papers and also started from a subject-based linked 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 self) with date of latest follow-up to facilitate tracking of responses. This would provide a much more flexible format than usenet, specifically avoiding awkward protocols for group creation and removal, and as well 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. Multi-component 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 – i.e. 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 such as doctorate or doctoral candidate, etc.; and permission to post/authorize 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 to motivate 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 usage of electronic networking 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 will make aggressive use of them if they are set up properly. It is anticipated that such systems will grow and evolve rapidly in the next few years.
- From such experimental systems, we have learned that open (i.e., 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 improvement expected in the near future.

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

1) Who will ultimately be the prime beneficiaries of electronic research communication (i.e., researchers, publishers, libraries, or other providers of network resources)?
2) What factors influence research communities in their rate and degree of acceptance of electronic technology, and what mechanisms are effective in facilitating such changes?
3) What role will be played by conventional peer-refereeing process in the electronic media, and how will it differ from field to field?
4) 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 whose primary virtue is instant retransmission?
5) What role will be played by library systems? (Will information be channelled somehow through libraries, or instead directly to researchers?)
6) 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 — 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?

7) 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, e.g. Acrobat or some other format currently under development, soon merge with the above or emerge as a new standard?)
8) 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, in which comments and errata dated and keyed to the relevant text can be inserted as electronic "post-it's" in the margins, etc. 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 artificial constraints of 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. (While it long remained a spare-time project with little financial or logistical support, as of 1 Mar 1995 it is supported by the U.S. National Science Foundation under Agreement No. 9413208.)

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 or professional societies, 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.

Paul Ginsparg 6/93, updated 11/93, 5/94, 4/95
e-mail: ginsparg@lanl.gov
MS-B285
Los Alamos National Laboratory
Los Alamos, NM 87545

Acknowledgments. Many people have contributed (consciously or otherwise) to the development of these systems. The original distribution list from which hep-th sprung in 1991 was assembled by Joanne Cohn, whose incipient efforts demonstrated that members of this 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 remote-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. Dave Forslund and Rick Luce helped lobby for support from within the Laboratory, and the Advanced Computing Laboratory has in addition provided some logistical and moral support. Finally, Geoffrey West repeatedly and against all obvious reason insisted that Los Alamos lab is appropriate to sponsor this activity, while simultaneously bearing the bad news both from within the laboratory and from certain government funding agencies.