A HISTORY OF THE BANACH SPACE ARCHIVE AND IMPLICATIONS FOR ELECTRONIC ARCHIVES OF PUBLICATIONS (PRELIMINARY VERSION)

DALE ALSPACH

The Banach space list and archive was begun in August 1989 and was founded using a list serving package on a DEC VMS VAX. I became aware of this type of software for the first time and realized that it had the potential to implement an idea for a Banach space newsletter that Pete Casazza had proposed. Pete had been discussing the idea of having a newsletter that appeared several times a year and would announce new results, advertise conferences, and provide a forum for queries. Because of the amount of work, startup costs, and other considerations Pete's idea was never implemented in paper form. The advent of email and list server software made it possible to implement some of Pete's ideas without the costs of production and of mailing a printed newsletter.

The initial list originated from electronic address lists that had been gathered from individual researchers in Banach space theory. At that time there were between 30 and 40 addresses that were used regularly and it was common to receive a message with all of the addresses listed. This list became the subscriber list at start up of the Banach archive. The list functioned by email only. I had no real experience with running such a list and no experience with being a subscriber to one. Thus the way the list functioned evolved slowly as I gained experience.

At first the list was unmoderated. Anyone who knew the address could send something to everyone on the list. Thus often papers that were in TeX or LaTeX were sent to all subscribers. For a time this was satisfactory, however several circumstances led to a change. One problem was the repeated accidental distribution of messages. In 1989 email was a new thing for most people and many of the people using it were not very familiar with computers. It seems to be an initiation rite for almost everyone who uses a list to confuse the list command address with the distribution address and unintentionally send a subscription

Date: January 8, 1999.

command to all of the list subscribers. This happened frequently in the early days of the Banach archive and I have noted that it still happens on many unmoderated lists today. This is an annoyance but there were also some accidental distributions of sensitive data. Finally some users complained that because of cost and/or limitations on mail inbox size that large mail messages were unwelcome. (The practice of charging for storage space or bytes transmitted no longer affects most of us but may return someday. [Metc, MM]) The list thus became moderated and whenever a paper was added to the archive, only the abstract was emailed to all subscribers.

As the number of papers added to the archive increased so did my workload and I began to look for ways to enhance the functionality of the archive and decrease the demands on me. In order to provide uniform information about each paper and to automate some of the procedures involved with posting papers, a required header for each paper was designed and instituted. This included obvious items such as the authors' names, the title, an abstract, a mathematics subject classification, TeX format, etc. One item that presently would seem unusual that was included on the header was a list of the standard printable characters and their decimal ASCII codes (32–126). At the time the Banach archive began there were some standardization problems with the translation of certain characters between machines. The principal problem was between EBCDIC machines such as many IBM machines and those following the ASCII standard. The problem was particularly disastrous for TeX files because braces were often incorrectly translated. By examining this part of the header one could quickly determine if the email had passed through an incorrect translation and thus was damaged.

```
%Special character check block
%32   space         33 ! exclam. pt.   34 " double quote  35 # sharp
%36 $ dollar        37 % percent       38 & ampersand     39 ' prime
%40 ( left paren.  41 ) rt. paren.    42 * asterisk      43 + plus
%44 , comma         45 - minus         46 . period        47 / divide
%58 : colon         59 ; semi-colon    60 < less than     61 = equal
%62 > greater than 63 ? question mark 64 @ at
%91 [ left bracket 92 \ backslash     93 ] right bracket 94 ^ caret
%95 _ underline    96 ` left single quote
%123 { left brace  124 | vertical bar 125 } right brace  126 ~ tilda
```

The moderation of the bulletin board prevented accidental mailings, but it had its own problems. One question that arose was what messages were appropriate to distribute to the subscribers to a Banach space list. At first I screened only for offensive material (of which there was essentially none). Early, almost anything that was of interest to the subscribers as people as well as mathematicians was distributed. Thus some political material such as a plea for aid for some mathematicians who had been detained for political reasons and some statements on ethnic problems in Yugoslavia was distributed.

After a time I began to require that messages have some relevancy to Banach space theory or to people who worked in the area. This new policy resulted from two changes. One, the Banach space archive had outgrown the small circle from which it started and second, by this time there were other outlets for papers in other fields and other lists and electronic discussion groups for other types of electronic correspondence. Personal announcements such as deaths, births of children, address changes, were and are still permitted. Advertisements of books, conferences, and other material related to the field are of course allowed. Academic job postings and certain other aspects of general interest to mathematicians are also permitted.

This narrowing still didn't prevent some controversy. Some individuals objected to my preventing their announcements from being sent because I felt that they were too far off topic. In another case I naively distributed information that was provided by someone other than the organizers of the conference about accommodations at the conference. The material included some negative comments about some of the arrangements and resulted in a somewhat heated reply from the conference organizers. I distributed part of the exchange to the subscribers but had to resort to editing to prevent personal remarks from being included.

Below is a chart showing the number of subscribers to the archive each year.

Year	89	90	91	92	93	94	95	96	97	98
Number	40	80	120	180	230	280	300	300	398	406

I have not tried to analyze the growth pattern in terms of events that might have affected the archive, but it seems clear that the first four or five years correspond to the increased use of electronic mail and increased awareness of the archive.

Two years ago a web interface was added to the Banach space archive. Through it papers, messages sent to subscribers, and links to other sites of interest were provided. Forms which allowed updating of address book entries and automatic generation of the header for papers were also implemented. Approximately one year ago David Morrison approached me about adding the Banach space archive to the Los Alamos archives. After some discussion agreement was reached and with some assistance from Greg Kuperberg the papers from the Banach space archive became available at xxx.lanl.gov on April 1, 1998. The web site at http://www.math.okstate.edu/˜alspach/banach/ is still being maintained, but all additions to the archive are just links to either the files at xxx.lanl.gov or the interface at front.math.ucdavis.edu. Because Banach spaces is not a separate designation at the xxx.lanl.gov archive, I have now assumed the role of a filter for postings in functional analysis and operator theory. Subscribers still receive email notices whenever something that I believe is of interest is posted at the Los Alamos archive, but other postings are not forwarded.

The changeover has not been entirely painless. Some users have had trouble adjusting to the use of gzipped files and uploading of new papers slowed to a trickle for a time. There has been renewed activity of late, but it is not yet clear to me that the users have adapted.

1. Ruminations on the Future of Electronic Archives

My experience with running an archive for many years has made me aware of several issues which will continue to influence the nature of and usefulness of electronic archives. I have attempted to formulate these as principles that any developer or maintainer of a mathematics preprint archive should keep in mind.

Users adapt slowly if at all. Anytime a change in the procedures for uploading or retrieving papers from the archive was implemented, there were complaints and adjustment periods before usage returned to its previous level. While I have largely embraced the new technology and therefore sometimes would think “everybody should be able to ...”, I learned to stop and focus on the computer skills of some individuals that I knew well and tried to design changes so that those individuals would be able to use the system.

Software and computer facilities available to users vary greatly. Some mathematicians have very limited computer facilities. The tremendous drop in hardware prices has helped, but a more serious issue is the configuration of the hardware and software. Many mathematicians have only with trepidation ventured on to the web and know virtually nothing about how the underlying software and hardware works. The result is that something which seems to me and perhaps you as a simple task such as a downloading a postscript viewer and installing it as a helper application for a browser, is greeted by others as a bewildering and impossible task. Left on their own such a user will probably give up. Many departments don’t have enough manpower either paid or volunteer to go around helping individuals configure software on their machines. Worldwide there is great variability in the availability of hardware and software.

Mathematicians want to do math. This is an obvious truism but it is directly related to the previous point. Some mathematicians have *actively* remained ignorant of computers. Learning about computers, how to use TeX, how to use email, how to use a browser, etc., all require time and energy and do not necessarily help one do mathematics. Consequently, some mathematicians have not. They rely on secretaries to type their papers and others to to handle anything to do with computers. If they use email or a browser at all, it is because everything has been packaged so that it requires no understanding other than that of which buttons to click on to start the application and make it function minimally.

Contributors are few. As you can see from the table above some 400 mathematicians subscribe to the Banach space bulletin board, yet in the nine years of operation less than 400 papers were included in the archive and the vast majority of those came from about 10% of the subscribers. I believe that there are several reasons for this.

- Electronic archives are not yet part of the culture.

  I think it is now fairly safe to say that TeX is part of the culture. Most thesis advisors expect that a student will be able to type his or her thesis in some dialect of TeX. On the other hand I doubt that most thesis advisors will suggest that their students add their papers to an electronic archive.
- Copyright concerns and protection of intellectual property discourage participation.

  At this time the status of a paper on an archive is unclear. Is it in some sense published and thereby is cite-able in priority claims? If the article is later published in a print or even electronic version what is the status of the preprint on the electronic archive. Should it be withdrawn?

Some authors choose to upload their work as soon as it is in a reasonable form. Others upload a paper at about the same time that it is submitted to a journal, and some only upload a preprint after the paper has been accepted. I have heard various arguments for each of these approaches. Some believe that the mathematical enterprise is best served by the freest flow of information. For example they may do this as a way to seek commentary so that they may revise the work. Others treat uploading to an archive as a kind of publication and thus feel that anything uploaded should be well polished. Still others see a competitive advantage in uploading late (or not at all) in that their results are advertised to others only after they have had months to work on further steps. Others are fearful that their ideas will be stolen and published without proper credit.

From these principles and reflection on some recent history, there are some straightforward conclusions that can be drawn.

- To be widely used the interface to an archive must be easy to use and not require the users to modify their systems in any significant way.
- The system itself must be accessible to those who have limited access to the internet. This includes both uploading and dowloading.
- The posting of a paper to an archive must be treated as an act of publication. It must be the case that if it is discovered that an author's work which was available on an electronic archive has been used by another without proper credit then the incident is handled just as it would if original author's work had been published in a print journal. I am not speaking in terms of legal remedies, but rather that this should be a part of the culture.

To get some perspective on the current situation consider the recent past. Twenty five years ago when I was beginning work on my thesis photocopies were very crude and the preprints that were circulated were often mimeographed or offset printed. Papers were typed on typewriters some of which had some ability to change fonts, e.g., IBM Selectrics had various balls. But changing font sizes was extremely limited and special symbols were often written in by hand or "rubbed off" a sheet onto the manuscript. The cost of producing and mailing these preprints was an important consideration. The quality of these preprints was far short of the published version of a paper and thus they were sometimes difficult to decipher. This also meant that reprints were in great demand for their readability, for mathematicians who did not have access to a journal and to provide a conveniently accessible copy of the paper. (Especially for those whose library did not allow journals to circulate.)

All of this sounds very primitive but, consider what the requirements this system placed on the mathematician himself. First a secretary probably handled the actual typing, so in terms of production of the paper the mathematician only had to produce a handwritten legible copy. The typing of the paper may have been a somewhat frustrating experience depending on the secretary's abilities but the mathematician's part was solely the proofreading. The other thing that the mathematician had to do was use the library and mail effectively to obtain reprints, distribute his own work and to keep up with the work of others. This allowed mathematicians to work in isolation with minimal equipment. His main need was access to a library and mail.

Consider the current situation. Having a preprint readily available for download has obvious advantages over the previous situation provided the technology is there to support it and the average mathematician is able to use the technology. Because of the widespread use of TeX it is now possible with the right technology to search for a paper, find it, download it, process it, and print it in a form not too inferior to that of a reprint in a few minutes. Thus for those with the technology and the access, this system provides a tremendous gain over the situation of twenty five years ago. At the same time this carries with it a danger of actually limiting access to a group of technologically advantaged individuals or outright failing to fulfill its promise because it discourages too many from using it.

Notice that the system of electronic preprints has a great deal of overhead. This is particularly true if the mathematician must essentially maintain his own hardware and software. To avoid some of these difficulties it is important that the technological requirements remain low. In the U.S. access to the worldwide web is considered standard but in other parts of the world telephone connections are not sufficient to support web access. Perhaps ten years from now this will not be an issue, but for the time being it seems important that access be possible by some low cost non-real-time means.

Currently most archives do not really limit what is uploaded. As a member of the editorial committee of a print journal, the *Proceedings of the American Mathematical Society*, I have observed that approximately 10% of the submissions contain serious errors. Because of the unrestricted nature of uploading I would guess that the rate is even higher for an archive such as Los Alamos and that many authors simply abandon their uploads. By this I mean that incorrect results are not withdrawn and uploaded versions are not replaced by revisions. Part of this no doubt follows from the role of the archives as a preprint server and not a repository of finalized work. Nevertheless without some sort of oversight a server may contain a large amount of worthless material. This overhead can negatively impact the operation of the server and, in particular, any scheme for implementing content-based searches.

Another dilemma for electronic archives is the problem of replacement by revisions. If an archive allows an author to replace a paper at will, there is a loss of history which may defeat any attempt to establish proper credit and priority. On the other hand it is a service to those who want to read the paper to have a version which is correct and up-to-date as possible. Thus there is a conflict between two uses of an archive. Which is most important: that a user be able to find the most up-to-date and correct information or that the historical record be intact? Can both be accomplished? The archive at Los Alamos, for example, currently keeps all versions that have been uploaded and makes them publicly available. (See http://xxx.lanl.gov/help/versions for the policy.)

For published papers there is the possibility of after publication revision of the electronic copy. Should this be permitted? Should a post-publication comments appendix with comments by the author be allowed? What about comments by others?

2. Ruminations on the future of mathematical publishing

While the advent of electronic journals will no doubt change publishing greatly, the temporary effects of this may mask some serious underlying issues in the publication of mathematical papers. Some have said that electronic storage will solve for the foreseeable future one problem which many libraries face: the problem of sufficient physical space to store and to provide access to library materials. With the possibility of reducing shelves of bound volumes to a few CD's or off site storage with electronic access a library could conceivably provide access electronically to enormous amounts of material.

While this seems like a reasonable possibility there is another problem that this may mask and the storage breakthrough may actually allow us to ignore a festering problem. Here is a joke that I once heard that points to a significant problem.

> Two mathematicians are discussing a paper which has been submitted to a journal. One is the editor and the other the referee for the paper.
>
> *Referee:* I finally finished reading that paper.
>
> *Editor:* Well, that makes at least two people who have read it.
>
> *Referee:* Oh, two? You mean you think the author read it?

While it is true that there has been a vast increase in the number of papers published it is also true that readership of papers is extremely low. This poses several dangers. Unexamined papers may contain errors, results are proved and reproved because the earlier results are not known, and the only evaluation of quality is done by the referee and editor at the time of submission. It has also added to the current crisis in journal subscriptions. Because library budgets are limited and cannot keep up with the growth in publication, libraries must cancel subscriptions and refuse to begin new subscriptions unless old ones are eliminated. The library at Oklahoma State attempted to track usage of non-circulating materials such as journals and use low usage as an argument for cancelling subscriptions.

Some, e.g., Kuperberg, Morrison, and Palais, [KMP], have heralded the electronic publication server as the solution to the library budget problem. Reasoning that much of the work in preparing articles for publication is done by the author and editor at little or no cost to the publisher, they see the potential to greatly expand the accessibility of mathematical work with most of the work performed by volunteers. Of course, the arrangement isn't really free, because costs are hidden in faculty salaries, the equipment used, the internet access, etc. The letter by Kuperberg, Morrison and Palais and an another article in the *Notices*, [BC], also touched on the role of commercial publishers in academic publishing and recently elicited a response by Edwin Beschler, [Bech], which addressed some of the issues from a different perspective.

Addressing the problem of low readership I believe will require a careful study of the nature and purpose of mathematical publishing. Is low readership really a problem from the mathematicians perspective or is it a problem because of its financial implications? Mathematics depends on the validity of results, but how many times must an argument be read before a result can be used with confidence? Some say that you should never use a result that you have not checked yourself. This may have been a reasonable position to take at one time but is it still? What about papers which contain laborious calculations which must be done to verify a result, but are not of themselves of much interest? Should such things be published? If so, where and in what form?

I believe that part of the difficulty with publishing mathematics and probably other academic publishing is that it does not fit the model of the rest of the publishing industry. As indicated above some mathematics is published solely as a notice that the results are valid and to make the evidence available should anyone care to look at it. The paper may not really be meant to be read by many mathematicians, but its existence is important. The commercial value of such a thing is small since few will want to buy a copy, but it is valuable to the mathematical community. Such publications differ from those whose purpose is to instruct or reveal new techniques and are thus really intended to be read. These may have some commercial value.

Let me end with what I see as one of the biggest issues concerning the growth of publication to be addressed. and where electronic archives could provide help. Suppose that you have just proved a new result but it is outside your area so you are unsure if it is known or not. How do you search the literature to determine if the result is in fact known or not? If only abstracts or reviews can be searched as in MathSciNet it is very likely that a result which is not the main one in a paper will not be mentioned in a review or abstract. On the other hand what can one search for in the body of a paper? Can a search based on terminology only be successful? How is notation to be understood by a search engine? Only recently have search schemes become sophisticated enough to recognize that "brown dog" has the same content as "the dog is brown". With the potential to represent the same mathematical fact in thousands of ways how can a search and indexing system cope with this?

References

[Bech] Edwin F. Beschler, *Pricing of scientific publications: a commercial publisher's point of view*, Notices Amer. Math. Soc. 45 (1998), 1333–1343.

[BC] J.J. Branin and Mary Case, *Reforming scholarly publishing in the sciences*, Notices Amer. Math. Soc. 45 (1998), 475–486.

[KMP] G. Kuperberg, D. Morrison and R. Palais, *Mathematical journals should be electronic and freely accessible*, Notices Amer. Math. Soc. 45 (1998), 845.

[Metc] Bob Metcalf, *From the Ether,* January 20, 1997 Infoworld Vol 19 (3) Available on the Infoworld web site as bm012097.htm.

[MM] Jeffrey K. MacKie-Mason and Hal Varian, *The economics of the internet*, http://www.ddj.com/ddsbk/1994/1994_inf/mackie.htm

Oklahoma State University, Department of Mathematics, Stillwater, OK 74078

E-mail address: `alspach@math.okstate.edu`