EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

CERN – SL DIVISION

POST MORTEM OF
THE ELECTRONIC PUBLICATION OF
THE EPAC 2000 PROCEEDINGS

P. Le Roux, J. Poole, CERN, Geneva, Switzerland

Abstract

The proceedings of the Seventh European Particle Accelerator Conference were the third in the series to be published electronically. This report describes the preparations before the conference, the activities at the conference and the work afterwards which was required to produce the CD-ROM and HTML versions of the proceedings. The whole process was a great success, with the proceedings available on WWW and for the publisher very soon after the conference. There were some areas where procedures could be improved and these are described in this report and statistics concerning the various aspects of the editorial activities are included.

Geneva, Switzerland

April 27, 2001

Contents

1 Introduction

The proceedings of the Seventh European Particle Accelerator Conference were the third in the series to be published electronically. This report describes the preparations before the conference, the activities at the conference and the work afterwards which was required to produce the CD-ROM and HTML versions of the proceedings. The paper volumes were produced from the electronic files.

Since the previous conference there has been significant progress in a couple of areas concerning electronic publishing: there is new software and new procedures have been developed. However, the total number of man-months required to produce the proceedings was more than in the previous EPAC conferences, even allowing for there being $\sim$7% more papers in 2000 than in 1998.

The final version of the proceedings were available on WWW in less than eight weeks and the CD was delivered to the Vienna team a few days later. There were some areas where procedures could be improved and these are detailed in this report together with statistics concerning the various aspects of the editorial activities. The publication of the proceedings of previous conferences were described and analysed in detail in [1, 2].

2 Abstract Submission and Publication

For the first time an Oracle database was used for the submission of abstracts. The submission procedure was very similar to that of previous years using a web form to gather the data which was backed up in a file and entered directly into the database. Data was imported into FileMaker for internal use, where it was verified and prepared to serve as the database for the conference. Following what was done for PAC99 it was possible for authors to review abstracts and submit corrections. The whole process worked very smoothly and reliably and further use will be made of Oracle in future conferences.

3 Instructions for Authors and Website

The EPAC website was an updated version of the Stockholm pages and was quickly built. Changes in templates and experience gained at EPAC'98 and PAC'99 were reflected in the instructions to authors and in the help on electronic publication. The Website with links to the JACoW templates was ready for authors at the beginning of April 2000.

4 Templates

It was agreed at the JACo Workshop held at Brookhaven Laboratory in December 1999, that standard templates should be made available for all conferences and this was reflected by installing them on the JACoW site, rather than the individual conference websites. The EPAC templates were already updated along these lines for ICALEPCS'99 and successfully used by that conference.

Before the templates are made publicly available the network of support personnel in laboratories around the world were asked once again to test and review the templates and procedures. The fact that the templates are now relatively mature was reflected in the low number (two) of changes that were required. For the first time at EPAC, the templates were only available via the web - the FTP server was not used.

The two most recent versions of WORD for the PC were WORD6.0 (95) and WORD97 and on Macintosh, WORD6 and WORD98. The LaTeX 2_ε templates were updated in terms of content and format, but there was no change in the version of LaTeX. A total of 10 templates were therefore prepared and tested.

5 Electronic submission

Once again submission of papers by FTP ahead of the conference was encouraged. Authors had to use FTP (or Fetch etc.) by hand in order to send their contributions. A web form was used to channel information to accompany the FTP submission. The script associated with the web form was essentially the same as the previous version, checking that the fields have been filled and that appropriate values had been used. The script also sent E-mails to the editors announcing the arrival of the files and giving the data content. Finally a confirmation was given to the author via the web browser and in an E-mail (the latter checked the validity of the E-mail address supplied by the author).

The FTP server was a UNIX server located at HEPHY, Vienna [1]. 12 Gbytes of mirrored storage were available and $\sim$10 were used by the end of the processing (before compression). The server was available from the time that submissions opened (early April) until well after the conference. After the conference it was used for authors to make their re-submission and to exchange files with CERN.

The reasons for having FTP submission ahead of the conference are to reduce the volume of work at the conference, to allow problems to be fixed and to give feedback to the authors before they leave their home institutes. About 35% of the papers were submitted ahead of the conference but none of them were processed before the conference. Without the processing there is little point in asking for FTP submissions before the conference. The only advantage was that some information could be inserted in the database ahead of the conference and this should be automated for the next EPAC conference.

The facility to submit a postscript file for distillation was not available and there were no comments about this

---

[1] For updated information and data originating from HEPHY relating to this report, please see the report from W. Mitaroff made at the JACoW team meeting in Frascati, April 2001 (http://cern.ch/JacoW/TM-2001/).

and it did not have any noticeable effect on the overall quality of submissions. One can conclude that the process is sufficiently mature that this procedure is no longer needed.

Authors were requested to bring a paper copy of FTP submitted contributions to the conference - this is essential for the quality checks.

6 Resources for the Proceedings Office

6.1 At the Conference

The factors which determine the resources required by the proceedings office remain the same and are:

- the number of papers to be processed
- the need to process the postscript on an appropriate platform - Macintosh files should be distilled on a Macintosh and PC or UNIX generated postscript on a PC and therefore the relative split MAC/PC+UNIX is important.
- all files should be accessible on all platforms (some form of networked file system)
- adequate software installations

The hardware requirements for 2000 were based on EPAC'98 experience where about 22% of the papers were prepared on Macintosh. It has been the experience at every conference that it takes longer to process papers on a Mac and therefore some additional weighting has to be applied in calculating the relative number of Mac's. In Vienna there were 16 PC's and 8 Mac's. The computers were linked by a local intranet and the files networked across all platforms.

The majority of the Macintosh's did not have floppy readers and therefore one of the processing team was appointed as *floppy man* who was charged with transferring all of the files from the floppies to the server. A similar process was implemented for the PCs because they were written from an account with read-write privileges whereas the processing was done with read-only on the directory containing the originals.

Most of the machines were equipped with CD-ROM readers and all of the PC's had floppy drives. Some PC's had Iomega zip drives and one was equipped for 250 Mbyte diskettes. The server was connected to the academic internet through a couple of Cisco routers and used a 100 Mbit/s link.

Once again we were fortunate enough to have support from the Asian conference series in the form of one full time person and his personal computers and part time help from Yong Ho Chin. With their help it was possible to resolve most of the Asian font problems at the conference.

The software inventory was as follows:

PC

Norton Antivirus
Netscape, telnet and FTP
Microsoft Office2000 (mainly for Word and Excel but Powerpoint was useful for making notices)
Acrobat4 with PitStop plugin

WinZip
Adobe Illustrator (on a few machines)
Adobe Photoshop 5 (on a few machines)

Mac

Antivirus
Adobe Illustrator
Adobe Photoshop 5
Enfocus Tailor
Microsoft Office98
Acrobat4 with PitStop plugin
Netscape
StuffIt/Binhex

Linux

$\LaTeX 2_\varepsilon$ from TeXLive
emacs
Netscape
telnet, FTP
CD-ROM burner and software
dat for backup

There was strong support from the other accelerator conferences (APAC, Cyclotron, HEACC, ICALEPCS, LINAC and PAC) for the proceedings and processing offices. In all there were about 18 people assigned to the processing office and a number of them remained for the week following the conference to continue working at HEPHY.

7 Processing the Files

7.1 Activities at the Conference

The main aim of the activity at the conference was author feedback achieved through the boards with coloured stickers. One person worked full time on keeping the boards up-to-date.

The contributions were handed in at the proceedings office by the authors and some initial checks were made:

- checking completeness of the submission
- cross-checking the information in the database (title, co-authors)
- inserting the keywords and number of pages in the database.

As usual, the submitted documents and accompanying sheets were placed in a transparent plastic folder so that the the submission sheet faced out at the back and the processing sheet was added at the front. The folders were split into two piles, for MAC and PC, and passed on to *floppy man* for loading and he then handed them on for processing. The processing sheet was used to record who had worked on the paper in the proceedings office, the acceptability of the contribution, to record any problems and to keep track of the status (processed OK, bad - see author, author seen and actions pending etc.).

Processing the files involved distilling the postscript and then verifying:

- the fonts
- the margins
- overall quality

A directory had been set up for each paper and the files were copied there by *floppy man*. The person doing the processing dragged the postscript from the the directory and 'dropped' it on the distiller icon. On successful distillation the PDF automatically opened in Acrobat, the checks were made and the file was cropped. An innovation in Vienna was the use of the form grid in Acrobat to check the margins of the submitted document – this proved to be a very useful aid but it was still necessary to print the file to be sure of full conformity.

Once the files had been processed the folders were passed back to the proceedings office where the database was updated to indicate the success or failure of the processing. The folders were then taken for the stickers to be placed on the boards and finally the folders were filed in the processing office.

The well defined path which the documents followed allowed a file to be found relatively rapidly. On a few occasions it took some time to locate a file but this was no more common than in more complex systems where files are tracked in a database - both systems are equally susceptible to human error.

7.2 Post Conference Activities

The post conference activities are principally fixing the problem files, careful checking, page numbering and making the indices, table of contents and other pages for the wrapper (photos, copyright, prize winners, conference organisation, list of participants, titles and separators etc.). In the problem cases, authors were contacted by E-mail and reminded that they had two weeks after the conference in which to resubmit.

The files for the Web are prepared first and then the CD-ROM versions and finally the paper version is prepared. It is of fundamental importance that the information in the database is complete and consistent and a lot of effort is placed on double and triple checking these data.

Once all of the files have been processed and deemed 'OK' they have to be scrutinised very carefully. The following steps are required:

- Open the file with Acrobat Reader or Exchange, check for error messages on all pages.
- Check that the file displays correctly and in particular that the position of the various elements like graphics are correct.
- Print the file and check that the margins and fonts are correct.
- Compare the printed version with the author's original.
- Cross check the title, number of pages and list of authors with the database.

The same procedures as were employed for EPAC98 were used again. The files were checked and where necessary fixed/resubmitted in Vienna and the basic PDF files transferred to CERN.

The Author Index and Table of Contents are produced with scripts and some new tricks were employed on this occasion to create PDF files with links directly from the database (see the Annex, Section 13).

The final versions of the PDF files containing page numbers, conference title and the hidden fields were published on the web as soon as they had all been verified. The verification involves printing the whole set and checking once again the layout, complete sequence of page numbers (including blank but numbered pages between sections) and the front- and back-matter. Even following all of these checks one paper was published with badly placed diagrams on its third page.

8 Statistics

8.1 Manpower

More manpower was used in 2000 than in 1998 but this was largely due to overkill at and in the week following the conference.

Table 1: Manpower Resources for EPAC Proceedings in Man-weeks

	1998	2000
R & D	2	2
Planning	2	2
Build/maintain WWW pages	4	2
Author documentation	1	1
FTP services	2	2(?)
Abstract Processing	5	5
Conference infrastructure	2	2
Processing at the Conference	10	20
Post Conference Processing		
1st week	8	13
following weeks	4	13
at CERN	8	8
	48	70

8.2 Computer Platforms

The statistics presented here are based on a sample of about 60% of the total number of the papers which was taken at the conference. Post conference analysis was not available at the time of writing .

There was a shift away from Macintosh (22% in 98 and 12% in 2000) but the number of UNIX users (identified as LaTeX papers) seemed to have increased from 11%.

8.3 Software used by Authors

The distribution of software packages used by authors remains dominated by Microsoft Word and LaTeX. The per-

centage of LATEX users has remained constant whilst non-preferred software has disappeared (∼2 papers), see Fig. 1.
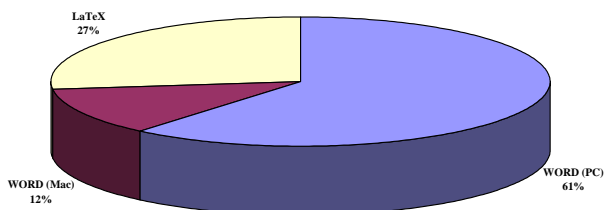

Figure 1: Software used for the preparation of EPAC2000 papers

8.4 Failure Rates

At the conference 15% of the postscript files submitted were found to have real problems and another 5% had small problems which were fixed by the team. In the weeks following the conference the papers were checked in more detail when this could be done in a more relaxed way and a significant number of papers previously thought to be OK were found to have problems. In general the failure rate was better than in previous years and is probably heading toward an asymptotic value of around 20%.

8.5 Fault Analysis

Nothing new in terms of problems was found in the papers processed in 2000. The same type of faults continue to be found with layout (margins) and font problems being the most common.

Some figures were very slow to display in PDF and these were fixed by saving them as bitmaps using Adobe Photoshop. However, a larger number of slow figures were accepted, reflecting the increase in computing power which is available now. It is possible that some people will have difficulties in displaying some of the papers in this year's proceedings and if there are a significant number of complaints, the acceptance criteria should be reviewed for 2002.

9 Publication

9.1 Preparing the CD-ROM

The files on the CD-ROM include all of the papers, the table of contents, the author index, the Acrobat index and the various other parts of the wrapper.

The structure of the CD-ROM was the same as in 1998, with only three files (Table of Contents, Author Index and Acrobat Index) in the root directory together with directories for Acrobat software, the wrapper, Acrobat index, the papers themselves and the whole of EPAC'98. The cover graphics and instruction booklet have also to be prepared. For 2000 the previous booklet was updated appropriately.

9.2 Preparing the Website

The website was built in exactly the same way as in previous years. The only major problem associated with this part of the process concerns the handling of special characters which have to be converted to HTML.

10 Problems Encountered in 2000

10.1 Printing

The PC's were running a version of Windows which most people were not familiar with and this gave rise to some problems when printing. The way in which to set up the printer from applications was not familiar to most people and as a result the parameters were wrong and it looked as if the margins were not correct in the papers.

There were a number of difficulties with the print service at the conference. The system blocked on numerous occasions causing delays and frustrations. This should be a simple matter to correct given a little time for testing.

10.2 LATEX installation

Once again the LATEX installation was not working at the beginning of the conference. In spite of explicit requests the TEXLive installation had not been used and no testing had been performed. After some considerable efforts the appropriate version was installed and made to function but two full days of LATEX processing were lost.

10.3 Floppy readers

It was a waste of expertise to have to allocate one person for Mac and another for PC to transfer files from floppy disks to the system. Whilst it is a good idea to place files in a place where they cannot be over-written this did not seem to be the case under all circumstances. There was certainly an initial period during which the files were not write protected and at a later stage there were certain operations across the network which allowed over-writing of original files.

10.4 Office materials

At the conference venue there are two offices to run - one for receipt of papers and the other for processing. A minimum of office materials are required - printer paper, pens, staplers, adhesive tape, waste bins and so on. None of these were available at the beginning of the week but they should have been.

The stickers for the author feedback should be available from the very beginning - they were not.

10.5 Networking and loss of files

A system of making the files available on Linux, PC and Mac was used (SAMBA) which gave rise to serious problems. A significant proportion of the processed files (PDF and corrected originals) were lost as a result of a bug in this system. These files all had to be re-checked in the week

following the conference. Files had been backed up and were recovered but one could not be sure if the backed up versions were in fact the "final" versions.

11 Improvements for the Future

11.1 FTP Submissions

An upload facility from the web browser has been implemented very nicely at PAC and this would improve the EPAC system - however, they did not have the FTP submission sheet at PAC'99. This can lead to problems if Mac and PC PostScript files are treated on the wrong platform.

A further enhancement in this area would be the capture of the information from the FTP submission sheet and automatic insertion in the database.

11.2 Floppy readers for Mac

It is inconvenient not to have floppy readers when we are asking for submissions on floppy. If the overhead is too great, then additional (less specialised) personnel should be allocated to perform the task of transferring the files to the file server. Perhaps this should be done as PAC'99 did - FTP only by the authors, even at the conference if they come with a floppy.

11.3 Computer Installation

Time should be allocated for testing the system - problems with print servers and software setup will then be identified ahead of the conference and not at a time when the stress is at a peak. Processing of FTP submissions would be a good way to check out the whole system.

11.4 LaTeX Installation

There is no reason why this cannot be done. The installation in Vienna was very good, once it was finally available. It only took about six hours to install and test the system once all of the components had been identified and prepared.

11.5 Processing FTP Submissions

FTP submission is becoming more and more popular - it should be better for everyone but if the files are not processed before the conference, the authors will be discouraged from making the effort.

11.6 Networking File Server

There was a near disaster in Vienna - a large fraction of the work could have been lost. It could have been even worse had the files been lost at a different time - if the loss was to happen during the working day before the backup there would be no backup and also the backup tapes were re-used later in the week which could also lead to problems. The area of file management and backup is mission critical.

11.7 Manpower for Mac Processing

The number of 'Mac' people in the processing office was much smaller than the number of machines available and therefore the Macintosh's were under-used. On the other hand there were more PC experts than machines. However, there was sufficient over-staffing that all of the Mac papers could be processed by Thursday afternoon anyway. Some effort is required to plan the staffing of the processing office so that new people can learn at the same timeas the core of experts is working and there should be even distribution of manpower throughout the conference.

11.8 Close Control over CD Production

A large effort went into preparing the files and final versions in the shortest possible time. It is very discouraging when the following stages of production are not treated with the same efficiency and determination. If the production of the CD is also handled locally to CERN and under the control of CERN/EPAC personnel the CD would be available in a very short time (posted within 12-14 weeks of the conference). However, this would cost an additional couple of man-weeks from this source.

12 Acknowledgements

References

[1] J. Poole, "Post Mortem of the Electronic Publication of the EPAC'96 Proceedings", CERN-SL-Note 96-68 DI, November 1996.

[2] L. Liljeby and J. Poole, "Post Mortem of the Electronic Publication of the EPAC'98 Proceedings", CERN-SL-Note 99-015 DI, February 1999.

13 Annex – New Procedures

Once the papers have been finalized, the files for the Web are prepared then the CD-ROM files and finally the paper copy version is printed.

The last steps of publication consist of multiple processes. The database is first updated with the final information (authors, titles, and PDF file names, number of pages per paper, keywords) and the next steps depend critically on the consistency of the data so double and triple checking is done. The papers are sorted into publication order and the page numbers are calculated with an SQL script, which also takes into account the blank pages and updates the database. Next, an SQL script which extracts the file names and concatenates them with a command line string is executed. This generates the "do_again.cmd" script that can be run on a Unix platform equipped with the Acrobat reader 4.x. For every PDF file, it starts the reader and converts the PDF into a PS file. Then, a PERL script ("add_page.pl") embeds the page number and the conference title in the PS file at the bottom of every page. Once this process has been performed for each PDF file, the resultant PS file is automatically distilled (distiller set up with a "watched folder"; all the PS files are output to that folder; and then wait a few minutes . . .).

The next step consists of inserting the hidden fields into every PDF file. To do so, another SQL script extracts data from the database (the file name, the title of the paper, the session, the concatenation of authors, and the concatenation of keywords). The result of the script is pasted in an Excel worksheet including a specific macro that has been reported elsewhere [2]. The macro is edited and the correct path to the PDF files folder is set up. Acrobat 4.x is started and then the macro is launched. It opens every file in Acrobat, embeds the hidden fields (from the worksheet) in the file, crops it (3mm on each side) and finally saves the file with the "optimised" option. Once all the files have been processed, the papers should be ready for publication. But, another check is required to verify that the page numbers, the conference title and the cropping are correct and the hidden fields embedded properly.

The Table of Contents (TOC) is the next task. Once again, an SQL script extracts the information required for the TOC (Title of the paper, file name, authors, page number) from the database. This information has HTML tags embedded so that the titles of the papers are formatted as hyperlinks with the file names as the URL. This automatically generated HTML code is pasted into a text file and the appropriate tags to create a valid Web page are added. At this point, the table of contents for the Web is ready. Then this page is edited with Word 97 (or later) to format the layout properly for the CD-ROM. Once, the page is well formatted, it is saved as HTML and then the "Acrobat PDFMaker 4.0 plug-in for Microsoft Word" (which should be available in Word if Acrobat 4.0 is installed) is used to distill the page into a PDF file. **The key feature of this process is that the PDF file still contains the hyperlinks.** At this stage the Table of Contents for the CD-ROM is also ready. The "Authors Index" and the "Keywords Index" (the latter for Chamonix workshops only) are created in the same way. An SQL script still extracts the information needed for each index (authors, file names, and page numbers for the "authors index" and keywords, file names, and page numbers for the "keywords index"). The information is gets HTML code embedded to make an HTML list sorted by authors or by keywords. The lists then have the appropriate tags added to make them complete Web pages. Finally, each Web file is formatted with Word to apply the layout required for the CD-ROM and then distilled in Word with the "Acrobat PDFMaker 4.0 plug-in" to obtain indices for the CD-ROM. For the paper copy, the same process is used except the SQL script doesn't mix the information with HTML tags, thus the output of the script is just a text file which can be formatted with Word.