

JACoW - A COLLABORATION SERVING THE ACCELERATOR COMMUNITY

J. Poole, C. Petit-Jean-Genaz, CERN, Geneva, Switzerland

Abstract

The Joint Accelerator Conferences Website started from an idea to publish particle accelerator conference proceedings on the WWW and has grown into an international collaboration supported by ten conference series. Through attendance at Steering Committee meetings and Team Meetings and through active participation in the work of the editorial teams of sister conferences, people with the responsibility for the production of the electronic versions of conference proceedings come together to learn from the experience of colleagues, and to develop common approaches to problems. The activities of the collaboration cover all aspects of electronic publication and have recently extended into conference scientific programme management. This paper reviews the history of the collaboration, describes some of the highlights in the activities during the life of the collaboration and presents the current status and future plans.

HISTORICAL DEVELOPMENT

In 1995 it was decided to publish EPAC's proceedings electronically and in preparation for this, the 1995 LEP Performance Workshop was used as pilot scheme. At the same time preparations were being made for the electronic publication of PAC'95 and EPAC was invited to collaborate in the processing in Dallas. It turned out that there was a very high failure rate in processing largely due to the lack of clear guidelines, author information and templates. This prompted us to develop a programme of author and editor education. This led to the formation of a team which went on to become the basis of the JACoW team. PAC and EPAC editors were invited to EPAC'96 and for the first time papers were processed at the conference to facilitate face-to-face feedback to authors at the conference.

Following the publication of EPAC'96 proceedings on the web, a Joint EPAC/PAC Website was proposed by PAC'99 Programme Chairman, Ilan Ben Zvi. The organisers of both series agreed to the principle and to give continued support to such a collaboration. APAC was subsequently invited to join and JACoW was formally set up after a meeting at PAC'97.

The JACoW team has always been involved in all stages of paper production from pre-conference instructions to paper submission, paper processing and finally to the publication of the proceedings. Attendance at regular meetings and sharing of skills was introduced as a formal requirement in the JACoW Terms of Reference in the year 2002, although it had been a feature of the *modus operandi* from the beginning.

The collaboration has grown steadily since that time and continues to attract new conferences each year. Although JACoW is based on an international collaboration in electronic publication of accelerator conference proceedings, it now does much more than just publish the proceedings as can be seen below.

WEBSITE FUNCTIONALITY

It was decided that the website should feature a search engine with a custom interface which would allow users to make Boolean searches across one or many of the conferences, by specifying titles, authors, keywords and/or full text searches. It was also decided that the papers must appear on the user's screen rapidly and with good visual quality. In order for the website to meet these requirements it is necessary for the PDF files of the papers to conform to certain standards. One of the first features introduced was a special paper size chosen so that the paper would print correctly centred on both US letter and A4 paper. This kind of constraint led the collaboration to set up templates for authors to use in the preparation of their papers.

The website provides a portal to information designed to assist authors in the preparation of papers for electronic publication, information about the collaboration and its meetings, the templates and of course the interface to the proceedings and the search engine.

Mirror sites were set up in USA and Asia with the aim of providing full functionality at high speed around the world. It proved to be difficult to provide the full functionality from the USA site but it was found that the performance from the European site was imperceptibly different, so the US site was abandoned but the Asian site was maintained. It was felt that users in USA were able to benefit from the huge bandwidths available to CERN whereas some of the Asian institutes were not so fortunate.

PAPER PROCESSING

Production of files for the website is never straightforward - there are always difficulties to make PDF files from some of the contributions and there are also a large number of papers submitted which do not conform to the JACoW specifications. The collaboration therefore has to train editors in the processing of papers and in the preparation of the PDF files for submission to the website. A cornerstone of author education is the famous dotting board, introduced at EPAC'96, which has been a feature of JACoW conferences ever since. A red dot against a paper informs the author that he should contact the proceedings office and we are able to



Figure 1: JACoW Team Members at the LBNL Meeting where the specifications for the SPS were drawn up.

explain the problem and hopefully how to avoid it in future.

The initial processing of a paper which has no problems takes around 15 minutes and results in a PDF file which is passed on to the final quality assurance process. At this final stage careful checks are made, the paper is compared to the original hard copy and the author list, title and other meta data are verified. The latter process takes around a further 15 minutes per paper and therefore there is a baseline investment of around 2.5 man-months for processing of the 'easy' papers for each EPAC. The papers which do not conform to the standard can take many hours or even days to fix - sometimes the authors are not capable of fixing the problems and the editors have to do the work for them. For a conference like EPAC, a total of around one man-year is required for the whole publication process.

TYPICAL PROBLEMS

Although there has been a huge improvement in the quality of the files submitted to JACoW conferences since the early days, there are still around 15% of the papers which cannot be successfully processed. There are many reasons for this (see Table 1) but usually it is because the authors have not followed the instructions. Although there are a small number of recidivists the red-dot system generally prevents authors doing the same thing again which helps to reduce the work load for future conferences. New software means that new templates have to be prepared and tested and it also means that new bugs have to be dealt with. The

Table 1: Distribution of problems encountered during paper processing at EPAC'02.

No PostScript submitted	8%
Format Problems	7%
Excessive Type 3 fonts (\LaTeX)	5%
Unusable files	3%
Multiple problems	3%
Font problems	3%
A4 printed on US letter paper	2%
Slow graphics	2%
Everything possible wrong	1%

constant *improvement* with new software versions and the arrival of first-time authors means that there will always be a minimum number of problem papers: this level is probably around 15% of contributions.

It is common practice for conferences to distinguish between problems which have been easy to fix and those which require re-submission or heavy modification. If an author does not submit a PostScript file it can be relatively easy for the editors to make one and produce a good PDF – in this instance the paper would probably get a yellow dot which signifies that the author should proof read the editor's version. This is because it is not uncommon that the original fonts used by the author are not available on the editor's computer resulting in small changes in such things

as equations or figures. The number of yellow dot papers is about the same as the red-dots but they only take a few minutes to fix. The figures in Table 1 correspond to both red and yellow dot papers.

Some of the most difficult problems to fix are those where an author has used some obscure software to produce a figure which results in a paper which cannot be processed to produce a PDF. In other cases where the margins have not been respected, an attempt to fix the problem can result in a paper which is too long and therefore the author has to eliminate some of the content. One of the most bizarre cases was a paper which had obviously been prepared with WORD but hand-written corrections had been made to the printed copy which was then scanned and the resulting PostScript file was submitted. At this conference the strangest case was where a paper was correctly prepared in every respect except that the PostScript file which was submitted was a mirror image.

An increasing number of authors have submitted PDF files to recent conferences in spite of instruction to the contrary. Editors are not able to use an author's PDF for a number of reasons – JACoW requires the PDF to conform to certain standards which means that the Acrobat parameters have to be set to ensure the correct behaviour for compression, fonts etc. The editors also have to process the PDF to insert data in the hidden fields, page numbers, conference banners and copyright information. All of these processes are done by scripts which rely on certain features in the PDF and therefore it is necessary for the editor to prepare the files using the correct parameters.

AUTOMATION IN PROCEEDINGS PRODUCTION

Data driven scripts have been used to facilitate some automation since the earliest JACoW publications. In the early days, this was only to produce html files, to insert the data in the hidden fields in the PDF files and to do the page numbering.

It is obvious that the scripts will not work if the data is not correct and for each conference a lot of effort goes into the verification and correction of the data, most of which is supplied by the authors. In order to produce an author's index one has to be sure that the person is correctly identified (the name must be spelled in the same way, with the correct initials and similarly his institute name). This has been impossible to achieve without at least 1.5 man-months per conference invested into the correction of the data entered by authors. It has been known for a single institute to be identified in about 10 different ways. In response to this situation it was proposed that a database containing author and institute data should be established so that this data could be used by the JACoW conferences and the idea for the Scientific Programme Management System (SPMS) was born.

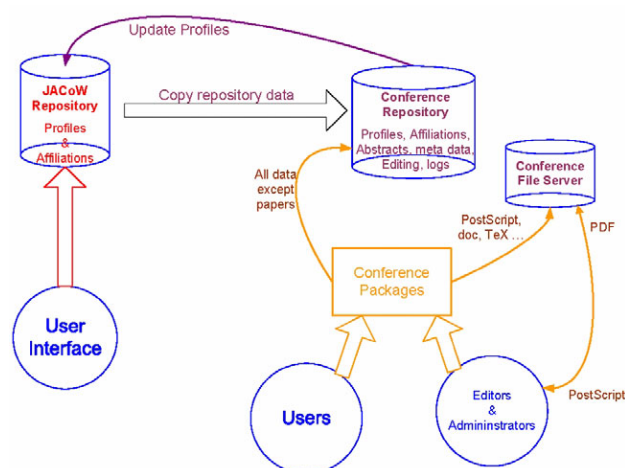


Figure 2: Overview of the Scientific Programme Management System.

THE JACOW SPMS

At a Team Meeting at LBNL in 2003, it was decided to investigate the possibility of producing a coherent package for the management of the whole scientific programme for JACoW conferences. A function specification and implementation plans were established.

There is a large volume of data associated with the scientific programme of any conference and in the past these data have been managed in many ways. During the 90's the trend was toward commercial database management systems and in particular Oracle™. Oracle has been used to manage parts of the data for many years and since PAC'99 it has played a significant role in PAC and EPAC conferences. In conferences leading up to 2002 it had been used to manage

- abstract data (abstract test, author lists etc.)
- programme data (paper_id, classifications, sorting order etc.)
- keywords
- editor comments and processing status
- page numbering

The SPMS is based on two pillars - a central repository of author profiles and institute data and a conference database system to be used by JACoW conferences. The conference systems are independently installed and they connect to the central repository so that authors do not need to enter their information more than once. Figure 2 shows how users can interact either through an instance of the database (the Conference Package for contributions and profiles/affiliations) or directly with the central repository (for profiles and affiliations). The conference repository will only hold the meta-data concerning contributions – the electronic files of contributions are uploaded to a separate fileserver.

Users are notorious for wanting their particular version of institute names and addresses and for this reason changes to

the way an affiliation details are defined are only permitted by the JACoW administrator. This ensures that the data are coherent and that author lists can be sorted as they should be, grouping people from one institute correctly. The way in which a conference uses a copy of the repository raises some consistency problems - if a change is made in one location, how is it replicated in the other? The solution adopted has been to automatically update profile data in the central repository when a user introduces a change via a conference database. Changes in the central repository are introduced in the conference copy through a scheduled synchronisation job, running at the conference installation.

The interface to the SPMS is web-based for both users and administrators and all of the code is stored in the database itself. The databases are customised through a set of parameters which are stored in the database and accessed through the web interface. A conference wishing to use the system downloads the system and only has to run an installation script to set up the system. The tuning is then done using the administrator interface.

MILESTONES

The following is a list of some of the major events marking JACoW's development.

- 1995 First electronic publications
- 1996 First website for proceedings
- 1997 First use of Oracle page numbers in PDF
- 1998 JACoW paper size
- 1999 Tracking of editor activities/comments
first Team Meeting/Workshop
- 2000 Cyclotrons, ICALEPCS and LINAC join
- 2001 Real templates for WORD introduced
Slides and videos included in proceedings
- 2002 Scanning project launched
JACoW domain name purchased
- 2003 SPMS created
- 2004 Award to JACoW
FEL, ICAP and RuPAC join

SCANNING PROJECT

Following an initiative by Gerry Jackson, the PAC'01 Programme Chairman, a project was launched in 2001 to scan proceedings from the pre-electronic era. An important feature of this project was that the resulting PDF files provide full-text searching capability. This is possible because there is a stage of character recognition in the production which allows most of the text to be identified. This proved so successful that EPAC has agreed to fund scanning of the early conferences in their series. This should mean that in the not-too-distant future, all PAC conferences (the first of which was in 1967) and all EPAC conferences (started in 1988) should be available on the JACoW site and be fully searchable.

THE FUTURE

JACoW has been providing a valuable service to the accelerator community for many years now and it can be said that it has reached maturity. In the coming years more conference series are expected to join (three applications have been approved at this conference) and the activities and services offered by the collaboration will extend. The SPMS has been a major step forward and in its next major release it is hoped to include a delegate registration package. This new package will make use of the information in the JACoW repository (making life simpler for users) and handle the fee payment.

ACKNOWLEDGEMENT

The success of JACoW is due to the dedication and good will of many people, too numerous to list here, who have contributed to the collaboration's activities since 1997. The recognition of JACoW's activities by PAC and EPAC is a great honour to be shared by everyone who has contributed over the years and we would like to add our own thanks to all who have been involved and we look forward many more years of collaboration and to an exciting future for JACoW.