



# Automated treatment of electronic resources in the Scientific Information Service at CERN

Nathalie Pignard (\*), Ingrid Geretschläger (\*\*), Jocelyne Jerdelet (\*\*\*)

## Abstract

We describe the automatic method of importation of meta data, developed in the Scientific Information Service, SIS, at CERN. The program, called *Uploader*, allows importation into the CERN library databases of bibliographic records and full text documents harvested from several Internet sources. The database sources offer essentially grey literature in physics and related subjects (e.g. DOE, KEK, Math-Doc, TipTop, etc.). This acquisitions policy is dependent on the automatic treatment of electronic resources and raises questions concerning the growing number of documents collected and on the enlargement of the subjects treated. Our constant efforts to enrich this meta data and to facilitate access to it, via the hyperlink model, brings new professional aspects to libraries.

**Keywords:** Grey literature - Automation - Document Importation - Electronic resources - Acquisitions policy

## From paper to electronic

For more than forty years, the Scientific Information Service CERN-SIS [1] has collaborated with research institutes and universities [2] world-wide to collect the work done by scientists.

CERN library regularly received, via mailing lists, papers from scientists of these institutes and universities. The documents, in paper form, were scanned to offer access through the Web to the users.

Today, this practice is diminishing and being transformed. Grey literature in science, particularly in physics, is available more and more in electronic form. Having distributed the documents for some years on both paper and electronically, many institutes have now chosen to use only the electronic route. This offers undeniable advantages over paper distribution: cost savings, quick and easy distribution, full text availability at a distance, the possibility to enrich the catalogue, cheap online access, etc. Maurice B. Line [3] points out other attractive aspects of the electronic document: "*Les principaux critères d'efficacité sont la rapidité de la fourniture du document, la fiabilité (la probabilité d'obtenir un document à partir de la ou des sources approchantes) et la facilité d'utilisation.*"

The virtual library has become a reality. Paper documents are more and more rare and the authors themselves generally prefer to submit their documents electronically. Also, most of the laboratories

offer access to their documents on the Web and have ceased to send out paper copies via mailing lists (Fermilab in the USA, Nordita in Denmark, etc. - signs and abbreviations are explained in the end of the article) and they encourage scientific libraries and researchers to consult their Web pages and databases.

Faced with this evolution, acquisition policies have to be reconsidered and adapted to the new standards of scientific information dissemination [4]. The Scientific Information Service CERN-SIS, and particularly the Document Management section, has progressively moved towards the automatic treatment of electronic resources. For some years, study and research projects have been regularly done on this subject by CERN-SIS [5], [6], [7], [8], [9].

The problem in this new context is the multiple consultation of databases. To find a document, a researcher has to consult many resources which is a time-consuming and boring task with often dubious results. To facilitate searching and to offer users a single search interface, CERN-SIS chose to import as many electronic documents as possible into the CERN databases [10].

In 1999, the informatic support team of the CERN library set up a program, the so-called *Uploader*, which allows automatic importation of bibliographic records extracted from several sources [11].

There are three main advantages with this tool

- the mailing lists sent out by institutes on paper have decreased but these papers can be found directly from the institutes' sites,
- it has increased the number of documents received from different laboratories and universities compared with the paper copies received before, one can also explore new databases offering documents of interest to the physicists at CERN and further enrich the library databases.

## **Automatic import of electronic records**

# **The functionality of the *Uploader***

From a file of data of any source (database or Web page), the program *Uploader* formats the records and adapts them to the cataloguing used in the CERN library databases [Annex 1].

For each source, configuration files are created to transform the original record to a record in the MARC (Machine Readable Cataloguing) format used in the CERN databases [12].

The program also has other functionality, e.g. updating existing records, searching for duplicates before importation, matching, etc.

## **The choice of sources**

The choice of which databases to explore was made according to several criteria. First of all, the Web sites of all institutes from whom CERN-SIS still received paper documents were consulted, to see if the institutes offered the same documents on-line. This analysis showed that more or less all institutes offer their publications on the Web, more or less elaborated. This study also pointed out that CERN-SIS received only a third of the documents available on the Web via mailing lists. There

are two possible explanations for this – it is likely that for economic reasons the laboratories make a selection of which documents to send out to other institutes; in addition mailing lists are not always kept up to date (profiles, address, etc.).

The need for automatic importation of these documents from the Web sites became obvious, but new problems arose, technical ones which we will comment on afterwards.

Other sources were explored, especially for subjects previously not developed much in the library databases. This is the case for mathematics (Math-Doc, Grenoble; mp\_arc, Austin, TX), or for theses in all subjects (e.g. Proquest [13], database hosted by Data Star).

Two methods to treat data found on the Internet

The sources explored can be divided into two types: Web pages and on-line databases. The functionality of each type is totally different and also therefore their automatic treatment by the *Uploader*.

#### *Web pages of research institutes*

Medium size laboratories and institutes which do not offer on-line databases, generally offer Web pages presenting the work of their researchers (most often theses) on their web sites [14]. Searching is quite primitive, as there is no real search engine implemented. Normally, the records are sorted by type of documents (theses, preprints, etc.) sometimes also by year. The number of documents is often limited. For this reason it is not always worth creating a special configuration for each Web page, a manual submission of the documents with their full text is much quicker. A second argument is the fact that Web pages are very unstable which makes it difficult to set up configurations for automatic importation of the documents.

A very important task is the follow-up of the Web pages: how to be alerted whenever a new record is added? Alerting services for these sites are rare. Only two sources propose this service: TipTop [15] (I.O.P, Bristol) for conference announcements and mp\_arc for preprints in mathematics. Another solution was to put alerts on the Web pages and to be informed when they changed. Around eighty alerts were placed on Web pages of some thirty institutes [Annex 2].

#### *Databases*

On-line databases often offer the possibility to do multi-criteria searching. In contrast to the web pages described above though, it is generally impossible to put an alert on the actual search results. It is therefore very difficult to import regularly by small periods (e.g. weekly) new records added to the database, except for those bases which offer an alert service.

If no alert service is offered, the method adopted is an annual search in the database for the previous year. This however means there is a delay of some months before receiving the bibliographic notices.

Another obstacle is the format of the search results. The majority of the time, they are displayed in a short list, with hypertext links to see the complete notice (e.g. first author only in DOE, *Department of Energy* [Annex 3.1], truncated title in CITHER [Annex 3.2]). In these cases, proper importation of notices becomes extremely difficult, if not impossible.

Also, within the same source, the cataloguing is specific to each type of document. Therefore different configurations are needed for each type of document in each database, e.g. FERMILAB preprints and theses. From July to December 2000, we wrote 14 configurations for 9 databases.

## Problems

### Instability of the Web pages

The Web pages are affected by several types of inconsistencies.

*Chronological instability.* Pages can disappear at any time, which is annoying if the URL imported just links to the bibliographic notice of the paper on the institute's website. Instead of the full text, the user will get the message "error 404". Therefore, the URL of the place where the full text is actually stored in the institute is also imported and stored on the CERN server whenever possible.

*Inconsistency in the structure of the pages.* For many configurations the html tags in the source file of the web pages allows the easy separation of the fields and sub fields in the bibliographic notices. However, the tags used are not always the same from one page to another, for different types of documents or even in the same Web page. In effect, in the majority of cases the pages are presented as free text and there is no common structure (spaces, tabulators, paragraphs, etc.) between the bibliographic notices. The constraints imposed by databases are non existent and there is no way we are able to write a configuration to import the notices. On the contrary, such notices have to be input manually.

*Inconsistencies in the bibliographic fields, not always catalogued according to the rules.* The main reason is that the institute's Web pages are not done by information officers, but by administrative assistants with no training in basic information science, causing heterogeneity in the bibliographic fields, most frequently – and most annoying, in the author fields (e.g. mp\_arc, Austin, [[Annex 4](#)]). Normally there is some coherence between the notices on the same page, i.e. author, title, number.

Other databases allow external persons to submit documents and create bibliographic notices (e.g.. TipTop for conference announcements and Los Alamos [16], which only accepts submissions from authors). This results in many irregularities and a complete loss of homogeneity in the presentation of the documents. Very often the information is presented in multiple variations, e.g. preprint numbers IUAP-00-xxx (number not yet attributed), CERN-TH-2K-1 (instead of CERN-TH-2000-1), MPS15600 (instead of MPS-2000-156), or the information is missing, e.g. the full author's list of a collaboration.

### Manual checking is still needed

These inconsistencies in the Web pages are not compatible with the rigid cataloguing structure required by the library catalogue. One of the main aims of the CERN information service is to offer users a coherent and homogenous database so they can obtain exact search results. For this reason manual checking, proof-reading and validation of the imported notices is necessary and CERN SIS continues to do this.

There is no doubt that the use of the Uploader programme offers considerable time saving, compared to manual submissions. It has also greatly increased the number of documents made accessible and available by CERN-SIS (see statistics for year 2000 in [Annex 6](#)).

However, for these procedures we need to select databases in accordance with the CERN research programmes, to study the layout of the bibliographic notices, to carefully implement working configurations, to import only the notices we want (avoid duplicates, non relevant subjects etc.) and to correct the notices (presented in a UNIX file, correction in Emacs or vi) before their validation and importation (corrections in Aleph have to be made notice per notice and are much heavier to handle). The richer the databases, the more time consuming the procedure becomes.

In addition, the instability of the Web pages requires a very close follow-up of the sources and constant updates to the configuration files. We conclude that with the Uploader tool, the work of the librarian changes (from manual submissions to automatic importation) but remains essential.

This evolution in the activities of CERN-SIS activity is part of the desire to add value to the bibliographic notices and to the search platform (WebLib2) to have a richer database and facilitate access for the library users.

## **CERN-SIS added value**

The added value offered by the library includes the correction of imported notices, updates to bibliographic fields and the addition of hypertext links between different kinds of information. [[Annex 5](#)].

### **Links between notices**

On the web version of the CERN-SIS databases, individual contributions to a conference are linked to each other and to the proceedings of the conference. With only one click the user, coming from an article, can access all the articles of the conference, the proceedings and/or the conference homepage, if there is one. The link is dynamic which means that any corrections will be transmitted to all notices linked together. A bibliographic entry for an article could link to the conference, to a journal and to the preprint. If the journal and/or the preprint are available electronically, the full text is then available to the user. A single notice can contain multiple links.

The link management has to be as safe and precise as possible otherwise the linking will not work.

### **Uniformity and standardisation**

Uniformity and standardisation of bibliographic fields is an important task. For example, standardising author names, including transliteration from Cyrillic and treatment of special accents (Russian or Nordic names) [[17](#)]. The goal is that all publications of an author can be found in one search, by using a standard orthography.

Standardisation is also applied to publication references: abbreviation of journal titles according to the standard ISO 4. A file of cross-references detects the multiple forms of journal titles and transforms them to the uniform title. A unique uniform journal title guarantees that the link from the article to the published e-journal version in the Web databases works.

## Adding information

Some databases only accept a limited number of author names (i.e. some thirty for preprints in the *e-print archives* Los Alamos). CERN-SIS adds all missed author names by extracting them from the PostScript file of the full text. This option is especially important for big collaborations which may have more than 500 authors.

Other information not included in the original notice is added to the CERN notice, e.g documents from CERN experiments. By deduction, CERN-SIS adds the affiliation, the division and the accelerator.

## Are we allowed to import data?

Until today, we applied this kind of acquisitions policy on a test basis. First of all we wanted to find out if and how importation was technically possible, what the interest would be for CERN-SIS in terms of time saving and for the research community in term of database enrichment. Now the question of the legitimacy of this type of procedure has arisen. In fact, these importations can not take place in the dark, without alerting the laboratories concerned. If the testing period goes well, CERN-SIS officially informs the institutes and asks permission to import some of their bibliographic notices on an exchange or cost-oriented basis. CERN-SIS has already reached agreements with Cornell University, NY, Fermilab, IL and databases like Inspec [18] and FIZ.

CERN-SIS acknowledges the imported data by a note "*record from...*" which is also valuable for the providers.

## Conclusion

The Uploader allows the importation of electronic resources and corresponds to the objective of CERN-SIS to offer to the research community an exhaustive database in high energy physics and border subjects. In addition to CERN documents, papers from institutes performing fundamental research in physics and related fields, i.e. Dapnia, KEK, SLAC, etc. are offered.

The aim is to offer a "clean" database where the work of verifying and correcting the data ensures that quantity does not become a substitute for quality, a possible risk with this type of acquisition policy. The 'added value' provided by CERN-SIS is essential if importation is to be more than just the simple addition of data.

Today, more than 90% of the notices entered into the CERN database (these statistics cover CERN grey literature databases: preprints, articles, reports and theses) are imported or created electronically. Of this, only 3% is generated from the CERN EDS server, researchers and secretaries. The rest are generated by the importation procedures described in this article [Annex 6].

Generally speaking, this form of acquisition policy adopted by CERN-SIS is a way to make up for the lack of avant-garde discourse on the creation of union catalogues in grey literature. In fact for more than thirty years, the idea of creating such catalogues has been regularly discussed. Again today one of these projects is in the front line, the so-called *Open Archives Initiative*, in which CERN-SIS will participate [Annex 7].

Unfortunately, most of the time these projects are confronted by a variety of problems from the outset: technical problems (it is necessary to adopt common standards), the time factor and political non-priority. This is why, for the present, CERN-SIS finds other ways to offer users access to documents in high energy physics and to implement and continually review an acquisitions policy for scientific grey literature.

**Annex 1 :** [return to text](#)

**Example of import : database of the KEK institute**

\* The original notice (received from the database KISS - KEK Information Service System)

**199827167 KEK Preprint 98-167**

**Ohuchi, N.; Tsuchiya, K.; Ogitsu, T.; Ajima, Y.; Qiu, M.; Yamamoto, A.; Shintomi, T.(KEK, Tsukuba)**

[\*Magnetic field measurements of a 1-m long model quadrupole magnet for the LHC interaction region\*](#)

[\[Scanned images\]\[The first page\]](#)

\* The notice formatted according to CERN-SIS needs

eng

1998

\$\$k 199827167

Magnetic Field Measurements Of A 1-m Long Model Quadrupole Magnet For The Lhc Interaction Region

Ohuchi, N

Tsuchiya, K

Ogitsu, T

Ajima, Y

Qiu, M

Yamamoto, A

Shintomi, T

\$\$n KEK \$\$p Tsukuba \$\$d Oct 1998 \$\$c mult. p

\$\$x [http://www-lib.kek.jp/cgi-bin/img\\_index?199827167](http://www-lib.kek.jp/cgi-bin/img_index?199827167) \$\$n Full text

KEK-Preprint-98-167

\* The notice on the CERN-SIS Web

**Magnetic Field Measurements Of A 1-m Long Model Quadrupole Magnet For The Lhc Interaction Region / [Ohuchi, N](#); [Tsuchiva, K](#); [Ogitsu, T](#); [Ajima, Y](#); [Qiu, M](#); [Yamamoto, A](#); [Shintomi, T](#);**

KEK-Preprint-98-167. - Tsukuba : KEK , Oct 1998. - mult. p. - [Fulltext](#) -

[Detailed record](#) - Mark record

## **Annex 2 : [return to text](#)**

### **Alerts and SDI services**

#### SDI services

Some sites propose SDI (Selective dissemination of information) services on the following pattern: on a regular basis, normally weekly, new bibliographic notices and sent by e-mail to those who subscribed to the service. I.O.P offers this possibility on their site Physics Web for conference announcements. Same principle for Mathematical Physics Archives (mp\_arc), handled by Austin University, TX.

This kind of diffusion list can be combined with other services: a profile is set up (search equation keywords, type of documents, periodicity). The profile search is done automatically daily or weekly and the results are sent by e-mail. It is normally possible to define the layout of the notices sent and to display the link to the full text. We wrote a configuration for these notices and import them with the Uploader. We use SDI profiles for data from FIZ and Inspec.

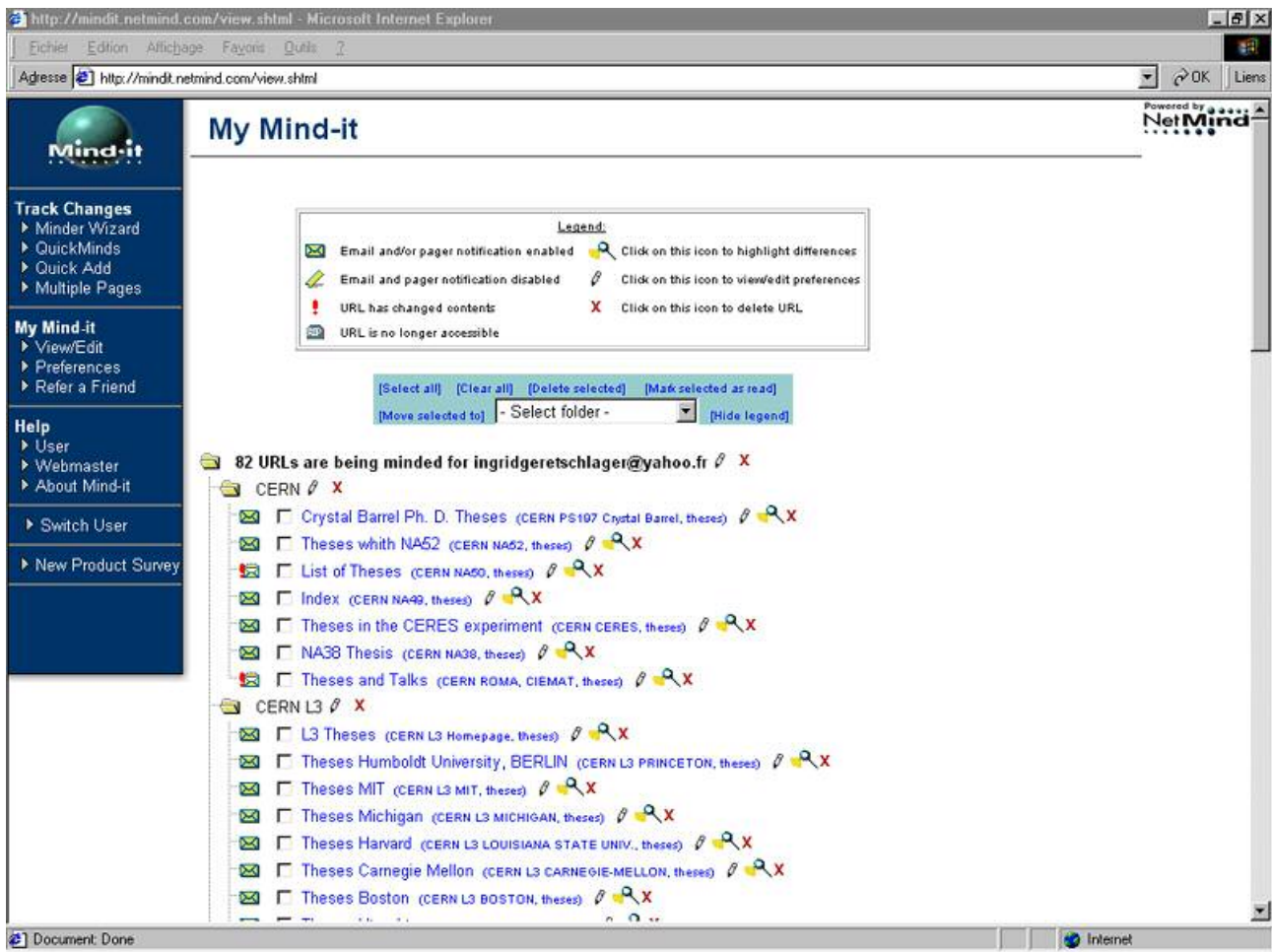
#### Alerts

For databases and pages which do not offer any SDI service, we set up alerts on Web pages we found interesting and which we think will grow. The alert is an automatic observation of the URL (uniform resource locator). We chose the free software Mind-It (MindIt / NetMind, <http://mindit.netmind.com/>). This tool regularly browses the URL addresses and detects all kind of change in the address and on the page: addition, corrections and suppression of data, migration of the address, closing of the page. Changes are shown by icons and highlighted in colour on the page. This is very convenient. Mind-It offers folders to help organises the alerts, to name each one and to define the periodicity of the browsing.

CERN-SIS tries to run Mind-It once a month and to submit the new bibliographic notices to the CERN EDS server. A manual submission is more time-consuming than a simple manual input, but with the advantage of transferring the full text file to the CERN EDS server for archiving. The CERN EDS server is stable so the file remains accessible.

When it is not possible to set up a profile for automatic importation an alert at least allows us to check on the evolution of the web pages and the publication of new documents.





**Example of alerts put up on the site Mind-It**

### **Annex 3 : [return to text](#)**

### **Examples of problems detected in sources when creating configuration profiles**

#### *Annex 3.1 - Example of a search result in the DOE database*

Only the first author is mentioned; to access the other authors and more bibliographic details, it is necessary to click on the hypertext link. It is only possibly to import incomplete short bibliographic notices.

http://www.osti.gov/bridge/search.results.jsp?queryId=6&start=1& - Microsoft Internet Explorer

Agresse <http://www.osti.gov/bridge/search.results.jsp?queryId=6&start=1&> OK Liens

## Search Results

Searched: ( Subject contains ( 74 ) and Publication Date > ( 19990101 ) and Publication Date < ( 20001231 ) sorted by Publication Date descending  
 Found: 19 matches, Results: 1-19 << Previous matches - Next matches >>

Primary ID	Title	Primary Author	Pub Date	Document
<a href="#">LA-UR-00-5736</a>	The QCD phase transition in relativistic heavy ion collisions -97-040	Cooper, F.	2000 Dec 01	 358K
<a href="#">PPPL-3623</a>	Modeling of Alcator C-Mod Divertor Baffling Experiments	D. P. Stotler	2000 Nov 29	 248K
<a href="#">SAND2000-2843C</a>	Synthesis and Characterization of Sn-Coated SFG10 Graphites as Negative Electrodes in Li-Ion Cells	VEERARAGHAVAN,B.	2000 Nov 15	 525K
<a href="#">WSRC-TR-2000-00298</a>	Characterization of Simulant LAW Envelope A, B, and C with Glass Formers	Hansen, E.K.	2000 Oct 17	 8001K
<a href="#">LBNL-45684</a>	A New 500-kV Ion Source Test Stand for HIF	Sangster, T.C.	2000 Oct 05	 26K
<a href="#">LA-UR-00-4910</a>	SOLVATION DYNAMICS OF ION PAIRS	Taylor, A.J.	2000 Oct 01	 1147K
<a href="#">LBNL-46787</a>	A 3.3 MJ, Rb(sup +1) Driver Design Based on an Integrated Systems Analysis	Meier, W.R.	2000 Sep 15	 43K
<a href="#">FFRMII AB-Conf00/207</a>	Beam-induced radiation in a muon storage ring for a	Nikolai V. Mokhov and	2000 Sep 14	

Terminé Internet

Annex 3.2 - Example of a search result in the CITHER database [return to text](#)

The title is truncated. To see the full title, it is necessary to click on the hypertext link. It is impossible to import the bibliographic notices.

http://csidoc.insa-lyon.fr/these/recherche.html - Microsoft Internet Explorer

Echier Edition Affichage Favoris Outils ?

Adresse http://csidoc.insa-lyon.fr/these/recherche.html

INSA-Lyon : Doc'INSA : CITHER :

Recherche de thèses par catalogue

Retour

Important : Ce formulaire vous permet de rechercher les notices bibliographiques des thèses de l'INSA de Lyon. Par défaut, seules les thèses électroniques sont sélectionnées (case cochée).

Cliquez sur le texte souligné pour visualiser les références complètes

1. 29 réponses

1999 | [ANTHIERENS Cedric](#) | [Conception d'un micro robot a actionneur](#) | C.83(2337),FC/2

1999 | [WALTER Helene](#) | [Modelisation 3D par elements finis du co](#) | C.83(2370),FC 2

1999 | [MANIN Lionel](#) | [Modeles de comportement multiniveaux pou](#) | C.83(2364),FC 2

1999 | [MARTINELLI Isabelle](#) | [Infiltration des eaux de ruissellement p](#) | C.83(2380)

1999 | [ZINCK Philippe](#) | [De la caracterisation micromecanique du](#) | C.83(2382),FC 2

1999 | [DEFAY Emmanuel](#) | [Elaboration et caracterisation de couche](#) | C.83(2381),FC 2

1999 | [PRIETO Flavio](#) | [Metrologie assistee par ordinateur](#) | C.83(2354),FC/2

1999 | [COLOMBINI Didier](#) | [Apports de la modelisation mecanique a l](#) | C.83(2361),FC 2

1999 | [LAFOREST Valerie](#) | [Technologies propres](#) | C.83(2341),FC/2

1999 | [DIAS Daniel](#) | [Renforcement du front de taille des tunn](#) | C.83(2355),FC 2

1999 | [GUIFFARD Benoit](#) | [Elaboration et caracterisation de cerami](#) | C.83(2356),FC/2

1999 | [PROST Thierry](#) | [Le risque, frontiere du genie urbain](#) | C.83(2353),FC/2

#### Annex 4 : [return to text](#)

#### Some author fields extracted from notices of mp\_arc

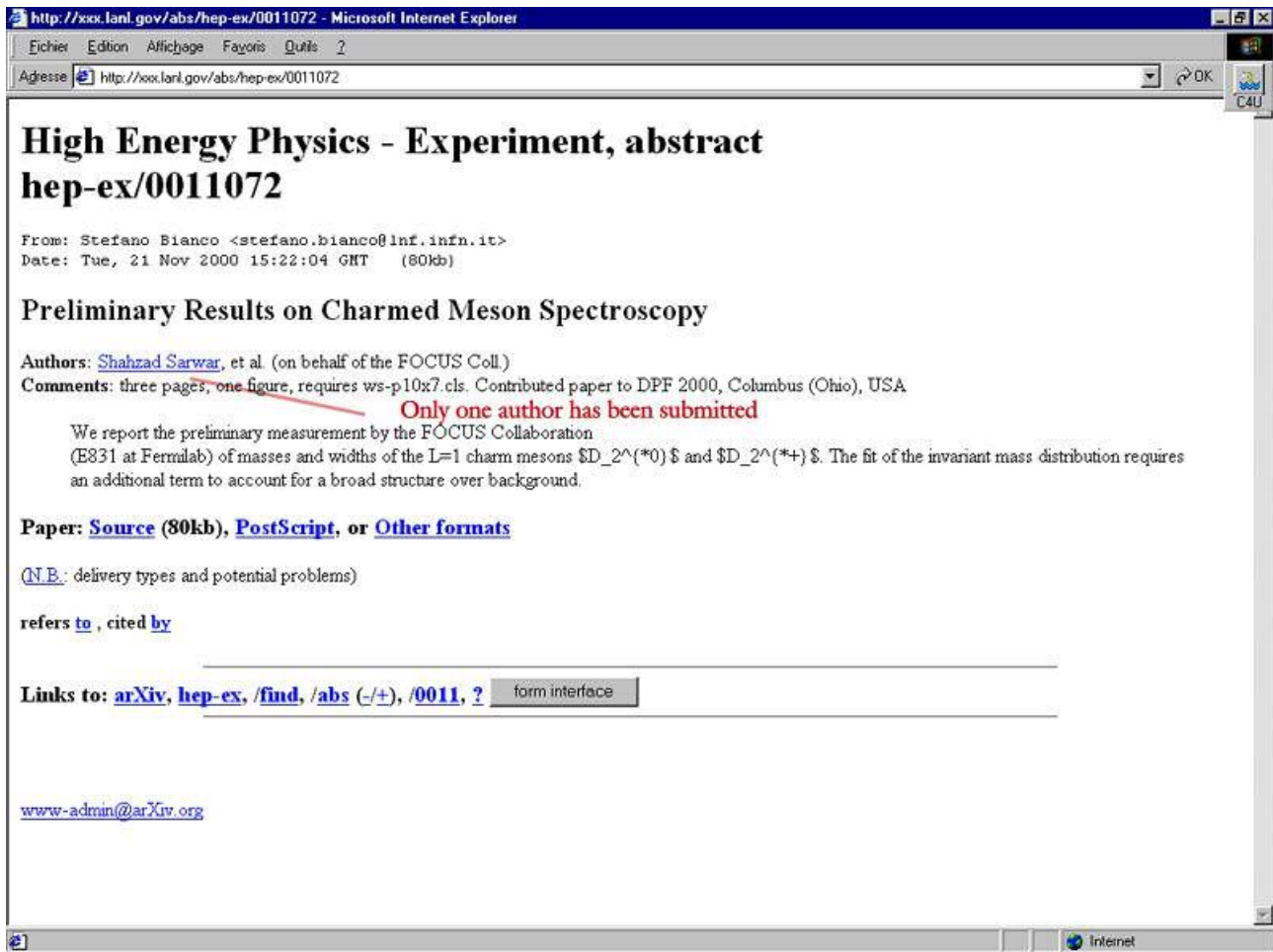
These author names were all entered in the base mp\_arc in the week from 19 au 26 October 2000. There are many inconsistencies which makes it difficult to create a good configuration for importation. Individual checking and probably manual corrections are mandatory for each bibliographic notice.

- Pavel Exner, Alain Joye
- A. Jorba
- J.Bricmont, A.Kupiainen, R.Lefevre
- Tai-PengTsai and Horng -Tzer Yau
- Werner Fischer, Hajo Leschke, Peter Mueller
- Bleher P., Ruiz J., Schonmann R.H., Shlosman S., Zagrebnov V.

Annex 5 : [return to text](#)

Added value : example of a notice imported from the preprint server in Los Alamos

Preprint submitted by the authors to the LANL Los Alamos server



The screenshot shows a Microsoft Internet Explorer browser window with the address bar displaying <http://xxx.lanl.gov/abs/hep-ex/0011072>. The main content area displays an email abstract for 'High Energy Physics - Experiment, abstract hep-ex/0011072'. The sender is Stefano Bianco <stefano.bianco@inf.infn.it> and the date is Tue, 21 Nov 2000 15:22:04 GMT (80kb). The subject is 'Preliminary Results on Charmed Meson Spectroscopy'. The authors are Shahzad Sarwar, et al. (on behalf of the FOCUS Coll.). The comments state: 'three pages, ~~one figure~~, requires ws-p10x7.cls. Contributed paper to DPF 2000, Columbus (Ohio), USA'. A red line is drawn through the text 'one figure'. Below this, it says 'Only one author has been submitted'. The abstract text reads: 'We report the preliminary measurement by the FOCUS Collaboration (E831 at Fermilab) of masses and widths of the L=1 charm mesons  $D_2^{(*)0}$  and  $D_2^{(*)+}$ . The fit of the invariant mass distribution requires an additional term to account for a broad structure over background.' Below the abstract, there are links for 'Source (80kb)', 'PostScript', and 'Other formats'. A note says '(N.B.: delivery types and potential problems)'. There is a section for 'refers to, cited by' which is currently empty. At the bottom, there are links to 'arXiv', 'hep-ex', 'find', 'abs (-/+)', and '0011, ?' with a 'form interface' button. The footer of the page includes the email address [www-admin@arXiv.org](mailto:www-admin@arXiv.org).

The same notice in the CERN-SIS database with added value

http://weblib.cern.ch/format/showfull?uid=0720959\_9343&base=CERCER&sysnb=2231638 - Microsoft Internet Explorer

Agresse http://weblib.cern.ch/format/showfull?uid=0720959\_9343&base=CERCER&sysnb=2231638

Preprint (hep-ex/0011072) [ http://weblib.cern.ch/abstract?hep-ex/0011072 ]

## Preliminary Results on Charmed Meson Spectroscopy

by: [Sarwar, S](#); Link, J M; Paolone, V S; Reyes, M; Anjos, J C; Yager, P M; Bediaga, I; Göbel, C; Magnin, J; De Miranda, J M; Pepe, I M; Dos Reis, A C; Simao, F; Carrillo, S; Casimiro, E; Mendez, H; Sánchez-Hernández, A; Uribe, C; Vasquez, F; Cinquini, L; Cumalat, J P; Ramirez, J E; O'Reilly, B; Vaandering, E W; Butler, J N; Cheung, H W K; Gaines, I; Garbincius, P H; Garren, L A; Gottschalk, E E; Gourlay, S A; Kasper, P H; Kreymer, A E; Kutschke, R; Bianco, S; Fabbri, F L; Sarwar, S; Cawfield, C; Kim, D Y; Park, K S; Rahimi, A; Gardner, R; Chung, Y S; Kang, J S; Ko, B R; Kwak, J W; Lee, K B; Myung, S S; Alimonti, G; Boschini, M; Brambilla, D; Caccianiga, B; Calandrino, A; D'Angelo, P; Di Corato, M; Dini, P; Giammarchi, M G; Inzani, P; Leveraro, F; Malvezzi, S; Menasce, D; Mezzadri, M; Milazzo, L; Moroni, L; Pedrini, D; Prezl, F; Rovere, M; Sala, A; Sala, S; Arena, V; Boca, G; Bonomi, G; Gianini, G; Liguori, G; Merlo, M; Pantea, D; Ratti, S P; Riccardi, C; Torre, P; Viola, L; Vitulo, P; Hernández, P; Lopez, A M; Mendez, L; Mirles, A; Montiel, E; Olaya, D; Quinones, J; Rivera, C; Zhang-Mayaguez, Y; Coptý, N; Purohit, M; Cho, K; Handler, T; Engh, D; Johns, W E; Hosack, M; Nehring, M S; Sales, M; Sheldon, P D; Stenson, K; Webster, M S; Sheaff, M; Kwon, Y;

All authors of the document are added

In: Meeting of the Division of Particles and Fields (DPF) of the American Physical Society (APS), Columbus, OH, USA, 9 - 12 Aug 2000 - [see corresponding record](#) - [List conference papers](#)

Conference information is added

- Email: Stefano Bianco [stefano.bianco@inf.infn.it](mailto:stefano.bianco@inf.infn.it)
- Collaborations: FOCUS Collaboration
- Imprint: 21 Nov 2000
- Pages: mult. p
- Subjects: Particle Physics - Experimental Results
- Experiment: FNAL E831 ([CERN Expt. Grey Book](#));
- [See record in SLAC SPIRES](#)

Link to the conference record and link to other contributions of this conference are created

Experiment information is added

**Abstract:**

We report the preliminary measurement by the FOCUS Collaboration (E831 at Fermilab) of masses and widths of the L=1 charm mesons  $D_{2^1}^*$  and  $D_{2^1}^{*+}$ . The fit of the invariant mass distribution requires an additional term to account for a broad structure over background.

Terminé Internet

Annex 6 : [return to text](#)

**Statistics : percentage of notices added manually or imported to the CERN database, between January and November 2000**

**Grey literature database :** articles, preprints, theses, reports

Total number of notices added to the grey literature database, from January to November 2000 = **approx. 53000**

Documents harvesting	Sources	Number of notices	Percentage
Manual input	Documents on paper or lists	4300	8%
Automatic import	CERN server (submissions by SIS, authors and secretaries)	1500	3%

	Los Alamos	29000	55%
	Others (INSPEC, SLAC, etc.)	4200	8%
	Tests done by SIS	14000	26%
	<b>Total import</b>	<b>48700</b>	<b>92%</b>
<b>Total of notices added to the base</b>		<b>53000</b>	<b>100%</b>

*Note : CERN-SIS database contains more than 350 000 notices*

**Annex 7 :** [return to text](#)

**The project *Open Archives Initiative*** (<http://www.openarchives.org/>)

The project *Open Archives Initiative* is a response to an appeal in July 1999 by Paul Ginsparg (initiator of the preprints database *e-Print archive* in Los Alamos), Rick Luce (LANL, Library) and Herbert Van de Sompel (LANL, Library). Their wish is for researchers and librarians in Europe and US to set up a universal service handling the auto-archiving of scientific publications by the authors.

The *Open Archives Initiative* has already resulted in conferences and some concrete proposals : The Santa Fe meeting (NM) on 21 and 22 October 1999, which gave birth to the " Santa Fe convention ", the workshop on 3 June 2000 in San Antonio, TX and another in September 2000 in Lisbon. The next *OAI* meeting will take place at CERN, from the 22 to 24 March 2001 [19].

The Santa Fe Convention [20] established a number of principles, particularly the recommendations for the implementation of interfaces allowing import of the meta data of each archive.

A site was created and a software allowing shared auto-archiving was developed by the department of IT, university of Southampton, England.

The goal of the *OAI* is that different libraries, by adopting common standards and a common so-called minimal notice, open access to their catalogues and offer easy exchange of data without heavy local modifications [21].

## Signs and abbreviations

<b>Institutes and research laboratories</b>	
DOE	U.S. Department of Energy, Washington, DC
Fermilab	Fermi National Accelerator Laboratory, Batavia, IL
KEK	High Energy Accelerator Research Organisation, Tsukuba, Japan
Nordita	Nordisk Institute for Teoretisk Fysik, Denmark
SLAC	Stanford Linear Accelerator, Stanford, CA
<b>Databases, projects ongoing</b>	
CITHER	Consultation en Texte Intégral des Thèses en Réseau, INSA de Lyon
FIZ	Fachinformations-Zentrum Physik, Karlsruhe
Inspec	Information Service in Physics, Electro-technology and Control
Math-Doc	Cellule de Co ordination Documentaire Nationale pour les Mathématiques, Univ. Grenoble 1
mp_arc	Mathematical Physics Archives, Texas Univ., Austin, TX

## References

[1] European Organization for Nuclear Research, Library <http://library.cern.ch/>

[2] i.e. GANIL (Grand Accélérateur National des Ions Lourds, Caen), DESY (Deutsches Elektronen Synchrotron, Hamburg), LAPP (Laboratoire d'Annecy-le-Vieux de la Physique des Particules, Annecy), MPI (Max Planck Institut, Garching), GSI (Gesellschaft für Schwerionenforschung, Darmstadt), RAL (Rutherford Appleton Laboratory, Chilton), DAPNIA (Département d'Astrophysique, de Physique des Particules, de Physique Nucléaire et de l'Instrumentation Associée, Saclay), SFB (Sonderforschungsbereich, Technische Univ. Berlin), Budker Institut for Nuclear Physics (Novosibirsk), Meisei Univ. (Tokyo), etc.

[3] Maurice B. Line. "Accéder ou acquérir, une véritable alternative pour les bibliothèques ?", *BBF* 41-1, 1996

[4] Isabelle Bontemps, Bernard Calenge (dir.). "Quelle politique documentaire pour l'acquisition de liens Internet en bibliothèque ?", Lyon : ENSSIB, 1999 : 67 p. Mémoire d'étude : D.C.B.  
<http://www.enssib.fr/bibliotheque/documents/dcb/bontemps.pdf>

[5] Isabelle Collignon, Ingrid Geretschläger (dir.). "Le traitement de la littérature grise à la bibliothèque du CERN", Geneva : CERN, 1998. DEUG-DIST : I.U.P./Univ. Lyon 1

[6] Catherine Deroche, Ingrid Geretschläger (dir.). "Automatisation partielle du traitement de la littérature grise dans le service d'information scientifique du CERN", Geneva : CERN, 1998. 59 p. D.E.S.S. Sci. Inf. : ENSSIB/Univ. Lyon 1  
<http://preprints.cern.ch/archive/electronic/cern/preprints/thesis/thesis-98-019.ps.gz>

[7] Catherine Cart, Ingrid Geretschläger. "Automatisation du traitement des documents CERN", 1999 : 6 p. *Soumis à* : Document Numérique  
<http://preprints.cern.ch/archive/electronic/cern/preprints/open/open-99-068.pdf>

[8] Philippe Ricanet, Jocelyne Milan (dir.), Ingrid Geretschläger (dir.). "Traitement de publications CERN de l'intranet : importation automatique/semi-automatique de publications d'expériences CERN dans le catalogue de la bibliothèque", Geneva : CERN, 1999 : 75 p. Maîtrise Documentation : Univ. Lyon 3 <http://documents.cern.ch/archive/electronic/cern/preprints/thesis/thesis-99-064.pdf>

[9] Nathalie Pignard, Ingrid Geretschläger (dir.), Jocelyne Jerdelet (dir.). "Comparative and statistical analysis between the CERN conference database and three other bases", Geneva : CERN, 1999 : 53 p. Maîtrise Information Communication : Univ. Lyon 2  
<http://preprints.cern.ch/archive/electronic/cern/preprints/thesis/thesis-99-060.pdf>

[10] <http://weplib.cern.ch/welcome.php>

[11] Martin Vesely, Jens Vigen (dir.). "Using Internet/Intranet Technologies in Library Automation", Geneva : CERN, 2000 : 67 p. Thèse : Univ. Economics Prague  
<http://documents.cern.ch/archive/electronic/cern/preprints/thesis/thesis-2000-040.pdf>

[12] Each "configuration" holds three main files. Two files allow to define the field structure of the original record, for extraction. The third file creates the new record with all bibliographic fields needed ; all sort of commands are applied to the original data to transform them to the needs of the CERN library catalogue.

[13] *Proquest Digital Dissertations* is a free but limited version of *Dissertation Abstracts International* (UMI). There are theses defences in North American universities 200 other universitites world wide. The period covered is the current plus the former year.

[14] Carole Clerc, Jean-Michel Mermet (dir.). "Contribution au développement d'un serveur de thèses électroniques", Lyon : INSA, 1999 : 72 p. Rapport de stage : DESSID  
<http://www.enssib.fr/bibliotheque/documents/dessid/clerc.pdf>

[15] TipTop, a *Unified Physics Resource* is the results of a private initiative of *TipTop* (Kenneth Holmlund, Mikko Karttunen and Günther Nowotny) and the database *PhysicsWeb* produced by IOP (Institute of Physics Publishing, Bristol). TipTop is maintained since 1998 by IOP to the attention of the research community in physics.

[16] arXiv.org e-Print archive / LANL, Los Alamos National Laboratory (Los Alamos, NM) since 1991 holds more than 170000 preprints and scientific communications in physics, mathematics and IT, before publication and offers the full text.

[17] i.e., Russian postfixes -ii, -ij, -y are unified to -y; forms ö, oe, o, Ø are transformed to Ø if relevant, etc.



[18] Inspec, the database produced by *The Institution of Electrical Engineers*, holds 7 millions of bibliographic notices since 1969. The base analyses most of the journals and proceedings in English in exact sciences.

[19] <http://documents.cern.ch/OAI>

[20] Herbert Von de Sompel et Carl Lagoze. "The Santa Fe Convention of the Open Archives Initiative", *D-Lib Magazine* 6-2, February 2000 <http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html>

[21] Librarians worldwide have defined since more than 50 years cataloguing rules and bibliographic notices layouts (minimal, standard, maximal) for the same purpose.

## Authors Details

Nathalie Pignard

Address : GRESEC, Université Stendhal, Institut de la Communication et des Médias  
Avenue du 8 Mai 1945  
F 38130 Échirolles

**Email :** <mailto:Nathalie.pignard@u-grenoble3.fr>

*doctoral student GRESEC (Groupe de Recherche sur les Enjeux de la Communication), actually at CERN*

**Ingrid Geretschläger**

**Address :** CERN  
ETT-SI-DM  
CH 1211 Geneva 23

**Email :** <mailto:Ingrid.geretschlager@cern.ch>

*Head of the Document Management section at CERN (European Organization for Nuclear Research)*

**Jocelyne Jerdelet**

**Address :** CERN  
ETT-SI-DM  
CH 1211 Geneva 23

**Email :** [jocelyne.jerdelet@cern.ch](mailto:jocelyne.jerdelet@cern.ch)

*Head of the Preprint unit at CERN (European Organization for Nuclear Research)*