

DESY-Library commission

Minutes of the 15th session from November 25, 2004, 11:00h in seminar room 1
Start: 11:00h, end: 13:15h

Present:

W. Drube (HasyLab), E. Fieder (ZEU-BIBL), A. Holtkamp (L), R.L. Johnson (Univ. HH), R. Klanner (FH), M. Köhler (L), W. Kreinacke (V), H. Mais (MPY), H. Münch (ZM1), H. Preißner (L), K. Sachs (L), D. Schmidt (L), F. Schrempp (T), F. Sefkow (FLC), R. Werner (EMBL)

Excused:

M. Behrens (IT), P. Folkerts (PR), C. Guse (ZEU-BIBL), T. Naumann (ZEU-Exp.)

Program

1. Opening of the session (R. Klanner)
2. Status report since the last meeting (M. Köhler)
3. Management flow within the Documentation (D. Schmidt)
4. Collaboration with SLAC (A. Holtkamp)
5. Comments about subject indexing (D. Schmidt)
6. Usage of the SPIRES-HEP data base (K. Sachs)
7. HEP literature retrieval - experimental groups (F. Sefkow)
8. Trends in the scientific literature search: theory (F. Schrempp)
9. Discussion

1 Opening of the session

Mr. Klanner opens the session and asks for approval of the minutes of the 14th session from June 9, 2004. The minutes were approved, there were no comments, some typos need correction. Mr. Klanner participates in this meeting for the last time as research director of DESY. His successor Mr. Heuer has not decided yet whether he will take over the chair of the library commission. A WWW page (<http://bibkommission.desy.de/>) is set up which links to the minutes of the meetings. At present the minutes are written in German and translated to English after approval. In the future the minutes will be written in English only. This proposal was accepted.

2 Status report since the last meeting

Presentation by M. Köhler.

2.1 Self check-out system

A machine will be installed for self check-out (Selbstaushleiherbuchung) of material from the library. It will be available 24/24, check-outs will be transferred to the system immediately and such be visible for other users. Renewal of the loan will be possible for every user via a web interface.

The system will be used with a user card. This card is necessary to borrow books. Outside opening hours (evening and weekend) the card works as key to the library. There was a limited call for tenders and the company Heutz from Niederkrüchten was selected. This is a small company with a lot of experience in self check-out systems, several libraries reported good reputation. The next step will be the installation and commissioning of the system. The machine will be delivered during December and should be in operation in spring.

Some discussion arose how to propagate the changes and necessary information to the user. A small note on the wall will not be sufficient. The main problem will be the introduction of the user cards. There won't be anonymous group cards for borrowing books. However, such cards might be possible as key-only version. Every user presently in the library system will receive his card automatically. The same card is valid as a key to enter the library outside opening hours and to borrow books. A write-up about the system will be propagated via the secretaries and the group managements.

Present status for the key functionality: it will work like other key cards at DESY. We have agreement from V1, hostel and PR; it is now with the works council (Betriebsrat). R. Klanner suggests, that the meeting of the 'Bereichsreferenten' should be contacted, to:

- comment and endorse the proposed scheme for library access and book rental,
- agree on the way the information is communicated to the DESY library users and the distribution of the cards is handled.

The card will have credit card format and the layout is approved by PR. It will show the name and ID-number of the user. The ID is necessary for the usage of the web interface as identification. The card will be valid for a limited period, the duration is stored in the library system.

For guests who stay at DESY for a short period (~14 days) a group card can be made available at the secretaries but as a key card only. A loan with anonymous group cards will not be possible to help tracking the books.

2.2 Electronic Library of Journals

All links on our web-page 'Electronic Journal' <http://www-library.desy.de/eljnl.html> now point directly to the EZB (Elektronische Zeitschriftenbibliothek Regensburg) journal data base, which is provided by the Regensburg University free of charge. It is a service to facilitate the use

of scholarly journals on the internet. It offers a fast, structured and unified interface to access full text articles online. At the moment, it contains 20515 titles, among them 2352 online-only journals, covering all subjects. 8009 journals can be read free-of-charge. Additionally, the participating libraries offer full text access to the respective journals they subscribe to.

Using the EZB gives our users a uniform interface to access full-text versions of journals and simplifies the support for the library.

2.3 ALEPH

Currently journals are managed via a KARDEX card-index, i.e. non-electronic which results for example in many user requests by phone. The use of the ALEPH system would bring several advantages for the user: transparency from the desk-top terminal, up-to-date information about new accessions, long-term loans online, manage routing lists, online booking and other HGF libraries can obtain information.

Advantages for the library: the management simplifies, information is available from every workplace, complaints will be delivered automatically, students can be trained with a state-of-the-art system, the manual KARDEX will be rendered unnecessary, better coordination with Hasylab library would be possible. The project is currently deferred since no position for a librarian is available. The introduction of the system would take about 15 month.

2.4 PoF HGF data base of Publications

Following a decision of the Directorate from January 2004 alternatives for a data base of publications meeting the PoF requirements were investigated. Candidates are mainly systems used by Hasylab and Jülich. Data models were developed and examined by the deputy directors (Bereichsreferenten) and B. Wittmann and J. Bonet. The needs for the development are being evaluated. There is a feasibility study for a HGF wide publication data base (meta search engine), too.

R. Klanner points out that the work on a HGF-wide data base for publication is probably substantial and that it should be pursued only with the consent of the DESY directorate.

3 Management flow within the Documentation

Presented by D. Schmidt.

In the last meeting the organization diagram of the library was discussed. Now the management flow of the documentation will be explained.

- Selection of literature; mainly done by scientists (Holtkamp, Köhler, Preißner); download is done by Mrs. Bierhahn.
- Classification, indexing: mainly done by 6 external free co-workers; about 19 000 documents/year; cross-check done by scientists of library (Holtkamp, Preißner, Sachs, Schmidt)
- Cataloging, input (Füllgraf, Lünert, Strate)

- Management of data base (mirror): done by scientists (Holtkamp, Köhler, Preißner); main work at SLAC; at DESY coordination of input for data base
- Maintenance of web pages, software management: done by scientists (Holtkamp, Köhler, Preißner, Sachs)
- Calendar of Conferences: together with SLAC (Holtkamp, Stein)

The electronic data base filled by the library comprises books (included in usual library catalogs) as well as journals, grey literature and new media, which are not included in usual library catalogs. The library selects contributions from text-books, monographs, conference proceedings, scientific journals, popular scientific magazines, preprints, reports and theses if they are relevant for HEP.

Workflow: selection (1 min), (short-)cataloging (5 min), proofreading (2-3 min) and correction (<1 min). After this step the article is available in SPIRES-HEP. Subject analysis of the document (classification, indexing) (~10 min), add keywords to the data base entry (1 min), proofreading (2 min), final correction (<1 min). There is a strong emphasis on the correctness of the information. Otherwise an article can not be found and the data base loses its significance. Numbers in brackets are estimated time consumption on average; amounts to a total of 22 minutes per article.

4 Collaboration with SLAC

Presented by A. Holtkamp.

SLAC is mainly responsible for the management of the data base, DESY concentrates on filling it.

A download of articles from some archives of arXiv is done automatically every day. Others like ‘physics’ are selected at DESY. Historically the emphasis at DESY is on journals. After the selection a cross-reference with the preprint entry in SPIRES is done semi-automatically (problems with changed titles and different articles with similar titles). If there is no preprint yet, the usual cataloging is done. This can not be done fully automatic since (unlike arXiv) every publisher has a different structure and a lot of effort is necessary to keep the required software up to date. For articles available on paper only the cataloging is done manually. The data input is always done at the SLAC data base, DESY hosts a mirror which also serves as backup.

Lately PACS numbers and free keywords (fk) as given by the authors are added to the data base. Up to now fk are available only for journals, they will be added for preprints shortly.

Other innovations are:

- a combined search (find k) in DESY keywords and fk. Later the search will include the title and will substitute the title search.
- a temporary entry of keywords before the final proof reading to speed up the process.
- conference talks of temporary interest won’t get keywords anymore.
- a new classification à la arXiv

- a new layout for the list of weekly accessions, which will in future include journals and conferences

R. Klanner: are there contacts to the publishers to facilitate the download of their information? – There are contacts to several publishers, e.g. SLAC to Elsevier and DESY to Springer. Are there commercial products?

R. Klanner proposes that an article (e.g. in the CERN Courier) should be written explaining the collaboration between DESY and SLAC on SPIRES.

5 Comments about subject indexing

Presented by D. Schmidt.

Experience with and comments from other data bases:

Dr. Holger Flachman (Universitäts und Landesbibliothek Münster)

‘Efficient subject indexing belongs to the very entity of cost-effective and quality-oriented libraries. This is especially true in the context of worldwide interconnected information systems.’

Manfred Wiencken (Informationsvermittlungsstelle des HMI, Berlin)

Dr. Luzin Weisel (Fachinformationszentrum Karlsruhe)

phone calls with both can be summarized as:

For literature searches on a specific scientific topic only a direct search using selected keywords leads to results with a high relevance. Query by keywords is the best navigation system in data bases with a large number of documents; it yields reliable search results.

Dr Luzian Weisel will talk on this subject (innerbetriebliche Fortbildung) in Spring 2006 ‘Ge-googelt und vergeigt - Informationsarbeit im Spannungsfeld von Kompetenz und Mittelmaß.’

6 Usage of the SPIRES-HEP data base

Presented by K. Sachs.

SPIRES log files at DESY were analyzed to find out how SPIRES-HEP is searched. The DESY mirror has about 80 000 requests/month, about 1 000 000 requests/month are at SLAC and there are further mirrors. Most searches are for exactly one article via reference or the author. Only 5% search in the title, 2% are in combination with the author. Conceptual searches are the remaining 3% title searches and 1% keyword searches. The small fraction of 4% corresponds to a large number of 100 conceptual requests daily at DESY and more than 1000 requests daily in total.

Conceptual searches via the title are not efficient since not enough information is available. Free author keywords are too various and not available for all articles. The abstract search is a good alternative next to the controlled keywords. The experience of other data bases (Dr. Luzian Weisel, FIZ Karlsruhe) is that normal users barely use more complicated terms than

1-word text searches. Professionals however, demand, value and use such possibilities. The host is responsible for the mediation of this information.

As an example 2 searches were compared:

- `find t w boson and structure function: result 1`
`find dk w and structure function: result 328`
- `find t muon cosmic rays spectrum: result 5`
`find t muon cosmic rays flux: result 0`
`find dk muon cosmic rays spectrum: result 1242`

If the keyword search gives too many results, the search can be refined.

The following information is available for conceptual searches: title, free keywords and abstract (in preparation) are included in the database within a few days after publication, DESY keywords follow later. There is a new search (`find k`) which is a word search over title, free keywords and DESY keywords. This means that with the same search term one gets some information immediately and can use the power of keywords. In addition variations of words (English vs US spelling, hyphenation, ...) are included automatically in the title search.

Up to now we have a large list of keywords which can be browsed. The best way to find a keyword might be to take a good article and look at the assigned keywords. We plan to sort the keywords by subject, improve the interface and introduce a web tutorial. Courses are an additional possibility. Further long term plans are to incorporate the abstract search into SPIRES-HEP. To facilitate the assignment of keywords we try to create proposals from free keywords, the title, abstract and body.

7 HEP literature retrieval - experimental groups

Presented by F. Sefkow.

Following a discussion at the last meeting we wanted to find out what (experimental) physicists search for, how they do it and what are the experiences.

Nine scientists were interviewed (mixed age; FLC / H1 / ZEUS; DESY / Uni HH / external).

What do you search: the search for articles dominates, for professors (teaching) books become important. Latest results are more important.

How do you search: for books a well sorted library is very important, with a call for improving the stock in the presence library. The OPAC catalogue was rarely mentioned. For articles SPIRES was named first and by all together with arXiv which has a search tool too. Young people named Google which searches the full document and finds all media, e.g. web pages which might lead to the wanted information. There is a new (beta) version Google Scholar (<http://scholar.google.com/>) for scientific information.

Most searches are for one specific article to get access to the online version. The community is small. In most cases the relevant person is known. One key article leads to others. Other searches use: author, title, journal, year.

Keywords are rarely used, if at all with bad experience: low efficiency, low purity.

8 Trends in the scientific literature search: theory

Presented by F. Schrempp.

After an introductory discussion with library persons 10 questions were sent to all theorists at DESY and II. Institut via email. From 90 people 71 replied.

- what age group are you? 28 student / 22 postdoc/ 21 staff
- did you know about the keyword project at DESY library? yes 35%, no 65%
- do you know SPIRES search title, date, author? yes 96%, no 4%
- how often do you use DESY keywords? frequent 8%, barely 13%, never 79%
- how often do you use SPIRES search criteria? frequent 94%, barely 4%, never 1%
- how often do you use free author keywords offered by journals? frequent 1%, barely 24%, never 75%
- how often do you use standard data bases (HEP, CONF, INST, EXPT, HEPNAMES, BOOKS)? frequent 75%, barely 18%, never 7%
- would you use DESY keywords more often after improvements of interface and documentation? yes 65%, no 35%
- do you think DESY keywords are useful to get started in a new field? yes 63%, no 37%
- do you prefer a faster 'mainstream' introduction of keywords? yes 30%, ? 48%, no 23%

Keywords are rarely used in all age groups, in general people who know and use keywords tend to be more experienced (staff). An improved interface would lead to a better usage. Several people pointed out that keywords are useful to get started in a new field.

9 Discussion

The keyword study was good and useful but the question came up whether there is a contradiction between the statements that keywords are useful and barely used? It was pointed out that the search in abstracts is very powerful and is possible at arXiv, Elsevier, Phys.Rev. and other journals. We should make an effort to get an abstract search incorporated in SPIRES, which might render the keyword search unnecessary.

It should be the responsibility of the journal referees to ensure a good title and abstract to improve the search and uniqueness of a paper. Sometimes keywords are too general. Understanding is needed to condense a paper into standardized keywords but this is not always enough to cover the contents. Abstracts are variable, several variations of a word can be used, which makes a search difficult. For most searches only a starting point is needed, e.g. the standard article of the dominant author. Other papers can be found via references and citations.

Currently we should continue with keywords. We have to think about how to continue in the long term.