Distributed Information Services in Physics

E. Hilf, M. Hohlfeld, T. Severiens, K. Zimmermann

27/03/2001

Abstract

The concept of distributed information services maintained by a distributed work force for scientific information is described. Realizations and experiences for Physics (since 1995), Marine Sciences, and dissertation theses in physics are presented. Technically the information is gathered from the local web servers of the worldwide distributed research institutes and departments, by distributed Harvest-gatherers, under the control of national learned societies or regional other institutions. Queries are answered by a network of mirrors of Harvest-brokers. For PhysNet, a Charter sets the rules to assure a noncommercial, free full-text access service, under the control of the set of national national learned societies involved, but unbiased to any single one.

Content

1. Concept

The web changes the usage of information management systems concerning creation, distribution, retrieval and archiving of scientific results.

In the former paper age the sequence of service steps was dictated by the technical requirements - and by tradition - from sending in a document written by its authors to a publisher, its refereeing, printing, distribution to libraries which registered and stored it there. The process was propelled financially by the libraries paying government money directly to the publishers, and with the reader/author not involved. This resulted in the well-described 'library crisis': a smaller percentage of published material can be provided to the individual reader by his/her local library. The reader/author is not involved in defining the requirements for new services that are possible in the digital age. The profits in the process go to publishers, not necessarily investing it into research and development.

In the digital age the individual steps of information management can be separated and optimized independently. This allows a real boost of innovation speed:

Posting a document on the web before refereeing triggers a priority time stamp and allows instant information of the worldwide scientific community to increase the speed of speeding.

Refereeing can be professionalized allowing parallel open as well as secret refereeing. Comments can be added by annotation.

Documents can be stored on any web-servers, and can even be stored and maintained at the author's server or that of his/her institute or department. This has the advantage that the author is normally most keen to keep the document in perfect readable shape, to correct errors, update it, and serve related material.

Retrieval can now be achieved by search engines without the former tedious work chain of librarian registration and catalogue mapping. This is achieved by the concept of machine-readable structural information on a document called metadata. In order to perform this task in a world of distributed archives and heterogeneous formats of documents, a universal standard of the definition of these metadata has to be set up. This has been achieved recently by the DublinCore metadata conference series.

Archiving documents for future generations is more and more seen as a task for national governments to be realized by the National Libraries. This huge task for scientific documents is still in an experimental status.

The professional *work force*, needed to propel all these services in the digital age using the web, can be distributed across the whole world.

Here we report on a set of services developed for the European Physical Society (EPS) to serve the Physics Community as a whole making use of the advantages described.

2. PhysNet - the Network of Physics Departments and Documents

Design

PhysNet is an information service for Physics, served by the EPS, which makes pragmatic use of the described advantages of the web:

Information provided by PhysNet is retrieved from the web servers of virtually all physics institutions, departments and institutes, and thus forms a large distributed database.

The quality and relevance to physics of the collected information is thus assured by the authors themselves, who are members of any of the professional physics institutions in the world.

A better retrieval by search engines used can be supported directly by the authors themselves. They add metadata to their documents.

The search engines used are based on Harvest, as developed by Colorado University. Several groups in the world are engaged in keeping this package of software up to date and work on its upgrading and further development [Cyclades (Dortmund), Uppsala, UKOLN, etc.].

Technically, Harvest allows the information to be gathered from the local institutional servers by a GATHERER, handing its index files to a BROKER which serves the incoming queries.

The advantage of the Harvest system is that it allows a net of independent but cooperating gatherers and brokers to be set up. This led to the setting up of a workforce distributed around the world.

National physical societies, regional centres or even local institutions can set up their own gatherer to feed their information into the global system. The advantage is that the institutions are thus able to steer and control the information they want to be fed into the system. They can allow or deny web-pages to be searched for by PhysNet. Think of local colloquia schedules evidently not of worldwide interest. At present about 27 gatherers are at work [link].

Analogously independent brokers are being set up by an increasing number of national physics societies.

Thus, PhysNet is designed to be seen as one service brought to the user by the national and international societies and their physics institutions.

Technical add-on services can be easily integrated into PhysNet by individual groups to the benefit of all. A spectacular example is certainly the ScientificTalk [link] by the Abdus Sallam International Centre for Theoretical Physics, Trieste which allows LaTeX formulae to be used in chatting.

The total workforce of PhysNet, by its distributed structure, is in principle infinite and grows steadily as more societies, institutions and authors are engaged.

EPS has set up a Charter to open PhysNet for all national and international societies to join and serve it. Also to assure that no single society can bias PhysNet. The input of information is open to all, but the quality is controlled by the EPS Action Committee of Publication and Scientific Communication.

In mathematics a similar distributed service has been set up in parallel and close cooperation, MathNet. It has been served by the German Mathematical Society (DMV) and is now adopted by the International Mathematical Union (IMU), controlled by its Committee on Electronic Information and Communication (CEIC). A Charter for MathNet, similar to the one of PhysNet has been designed, agreed on, and published by the IMU.

A very fruitful and close technical cooperation right from the beginning especially on metadata and retrieval has now been formalized by a cooperation agreement between the IMU and the EPS, representing PhysNet.

Set of Services

PhysDep [link] - Physics Departments Worldwide.
PhysDep allows one to search or surf across a set of lists of links to the servers of at present 1,760 institutions and departments of universities related to Physics, ordered by continent, country, and town. It is mostly the top two layers of pages on each server that is retrieved. In cases where local or regional gatherers have been set up, the search goes across the whole server, with the advantage that the local administrator can allow or deny specific documents to enter or to be withheld from PhysNet. This helps to sharpen the profile of the institution on the PhysNet.

PhysDoc [link] - Physics Documents Worldwide.
PhysDoc provides lists of links to document sources of worldwide distributed Physics Institutions. Such document sources may be preprints, research reports, annual reports, and lists of publications of local research groups and individual scientists. This complements central document collections such as journals, or the ArXiv. The lists are ordered by continent, country, and town. The service is complemented by a search engine which in addition allows one to search across the Mathematics Preprint Search System MPRESS. PACS and MSC classification can also be used to make it more diaphanous.

Documents where the author or his/her institution has added metadata according to the international DublinCore specification by just using the web-form MMM are well rewarded in that these documents are prominently retrieved.

The Journals site lists physics-related refereed, online journals which are available on the web for free (free access to full texts). A comprehensive list of the 'EPS Recognized Journals' is given as well.

At the Conferences site a collection of physics-related web servers is given, which lead to dates and meetings, workshops etc. They are maintained by various related societies, institutions, and service providers.

'PhysJobs' offers a list of links to various related job sites on the web. A Harvest-based search engine allows one to search for jobs at the listed services.

The Education service provides online educational resources for physics (e.g. Lecture Notes, Seminar Talks, Visualization and Demonstration Applets), listed by subject areas and partially sorted by level.

Links (to other Resources on the Web): This service offers a collection to further sources of Physics Information on the web. Links to online information services of other learned fields are also given.

By 'Services' useful online tools mostly concerning metadata are offered. By typing the information into easy-to-use web forms like MMM (My Meta Maker) the authors can enrich and improve their homepages and documents with metadata according to the international DublinCore standard.

An Upload-Interface is offered to allow the local institutions and authors to register their institution's and documents web pages, send additional URLs, update the entries or send messages to the PhysNet-Crew.

Societies, individual physicists, and physics institutions may join the PhysNet-Crew, for example by maintaining the link-list for their country or region or by setting up a local mirror of the PhysNet sites on their web server.

To improve the search results, institutions are recommended to install their own Harvestgatherer. They should maintain a link-list of their relevant local scientific document services by their individual groups.

Legal Aspects

All information of PhysNet is kept, stored, and maintained by the creators at their local institution's server. There is no centralized database set-up. Thus the creators of documents themselves retain all rights of their data and stay responsible for updating their own information.

PhysNet only gathers and processes the local information of physics institutions to make them accessible globally.

PhysNet is a noncommercial service. No commercials or banners are allowed. The access to information offered by PhysNet is free for anyone. The aim of PhysNet is to provide a long-term stable and distributed information service for physics with the collaboration of many national and international societies and physics organizations.

PhysNet is under the auspices of the European Physical Society (EPS) and several national societies. The Service is controlled by the EPS Action Committee on Publications and Scientific Communication (ACPuC). The technical development of services and standards are coordinated by the Institute for Science Networking at the Carl von Ossietzky University of Oldenburg.

Most recently PhysDoc has been made compliant to the Open Archive Initiative Protocol for Metadata Harvesting (OAI) which proposes an interface for document providers to service providers.

Since PhysDoc is not a single document provider but provides information from gathering several thousand local document providers which are at present not OAI compliant, special adaptions had to be made to the OAI interface to transfer the retrieved information to OAI service providers. The final step of registration of PhysDoc as one OAI document provider was taken on 25 February 2001.

Statistics and Usage [link]

To date, we have collected information from 1,760 physics departments worldwide. The number of the linked publication lists in PhysDoc seems to be stable at about 1,408 links, but certainly is still incomplete.

The number of documents reached by these links cannot be measured exactly, since authors may change their input without notice. Our estimate of the number of documents reached so far is well above 70,000. The number of stored links is 39,000 of which 988 (by February 2001) are offering DublinCore metadata.

It is a time-consuming work to keep the links for PhysDoc up to date, because links are often changed without notice by the authors at their local sites.

We have approximatily 800 requests per day. The service was installed in 1995.

3. Related Services in Learned Field and Research Projects

PhysDis

PhysDis is a subset of PhysDoc and focuses on Ph.D. theses and dissertations as a special type of publication in its dual role as exam work, prime scientific research work, and publication. In the past few years, the prestige of these documents in physics has changed from 'grey literature' to an important source dealing with the latest research results.

PhysDis currently offers 229 links to collections of Ph.D. theses and dissertations in 18 European countries, plus the collections of MIT and Fermilab in the US. We list 85 links in Germany, 48 in Spain, 21 in Sweden, 20 in Switzerland, and 55 links in other European countries. A total of 1,818 datasets including 250 full texts have been collected so far. The PhysDis service also uses a Harvest-based broker to allow for retrieval across all of the listed links. An upload interface is offered for the services to allow local institutions or the institutions' authors to register documents, information on documents, or lists of theses.

DissOnline

In the years 1998-2000 PhysDis was part of the DFG funded research project 'Dissertationen Online' in Germany. This interdisciplinary project of several learned societies involved chemistry, computer science, education, mathematics and physics, five German universities, computer centres, libraries and the German National Library (DDB). The main topics were retrieval, reading, printing, and archiving to reach a common work-flow and standard.

An agreement was reached with the German National Library concerning the set of metadata and the archiving of the electronic full texts. The results and recommended project tools can be found online.

OAD

A new German-US cooperation project 'Open Archives: Distributed services for physicists and graduate students (OAD)' is being launched simultaneously at Virginia Tech, USA and ISN Oldenburg, Germany. It is financed jointly by the National Science Foundation NSF and the Deutsche Forschungsgemeinschaft DFG. In this, PhysDoc will be coupled with the NDLTD archive (Networked Digital Library of Theses and Dissertations) in North America.

MareNet

In other fields MareNet was established for marine science. It contains more than 1,000 links to marine research institutions and documents worldwide, in detail to more than 450 marine research institutes and 220 documents. MareNet is under the auspices of the German Society for Marine Research (DGM - Deutsche Gesellschaft für Meeresforschung). The service was installed in November 2000 and now receives 150 requests per day.

4. Conclusion

PhysNet, PhysDis and MareNet are demonstrating that the web allows distributed heterogeneous databases, distributed work force resulting in one homogeneous service.

In the near future the more widespread usage of metadata will greatly improve the retrieval results for professional use in research.

Acknowledgement

5. References

1.   Thomas Severiens, Michael Hohlfeld, Kerstin Zimmermann, Eberhard R. Hilf
PhysDoc - A Distributed Network of Physics Institutions Documents Collecting, Indexing, and Searching High Quality Documents by using Harvest
D-Lib Magazine December 2000 Volume 6 Number 12 ISSN 1082-9873
URL: <doi://10.1045/december00-severiens>
2.   PhysNet URL: < http://physnet.uni-oldenburg.de/PhysNet/>
3.   PhysNet-Charter URL: <http://www.eps.org/PhysNet/charter.html>
4.   PhysDis URL: < http://elfikom.physik.uni-oldenburg.de/dissonline/PhysDis/dis_europe.html>
5.   Dissertationen Online URL: <http://www.dissonline.org/>
6.   Open Archives: Distributed services for physicists and graduate students (OAD) URL: <http://ins.uni-oldenburg.de/projects/OAD/>
7.   MareNet URL: <http://www.marenet.de/>
8.   MathNet URL: <http://www.mathnet.de/>
9.   Harvest URL: <http://www.tardis.ed.ac.uk/harvest/>
10.   OAI URL: <http://www.openarchives.org/>
11.   DDB URL: <http://www.ddb.de/index_e.htm>

Author Details

E. Hilf, M. Hohlfeld, T. Severiens, K. Zimmermann
Institute for Science Networking
Ammerländer Heerstraße 121
D - 26129 Oldenburg
Germany

Tel: +49 441 798 2742
Email: info@isn-oldenburg.de
URL: http://ins.uni-oldenburg.de/Institute/index_eng.html

E. Hilf Head of the institute, professor of Physics.

M. Hohlfeld, T. Severiens, K. Zimmermann researcher at the institute, physicists, involved in several information management projects (MareNet, PhysNet, DissOnline).