Scholarly Publishing and Information Choice on the Internet (インターネットにおける学術情報出版と情報選択)

Irene Langner (Universitäte Bonn)
Hiromichi Hashizume (National Center for Science Information Systems)

Abstract

Individual scholarly publishing on the Internet has constantly grown in recent years. This trend poses considerable challenges to the parties involved in traditional (journal type) academic publishing and gives rise to new problems of information selection and acknowledgement of academic achievements. In this paper we present a survey of different patterns of current scholarly publishing activities as well as patterns of information choice on the Internet. We show examples of how traditional players in the publishing process such as academic societies, publishers and libraries are trying to meet these challenges by engaging in electronic publishing activities themselves, and finally sketch a desirable future scenario of networked scholarly publishing, resource sharing and preserving, considering some possible (new) roles for authors, academic societies, digital libraries and various other kinds of intermediaries. What could be appropriate steps in such a direction?

Keywords:

electronic publishing, Internet, electronic journals, academic societies, digital libraries, quality selection, evaluation, search engines

1. Introduction

Printed scholarly journals have long been the main means of publishing research results in many academic fields. Partly because of the lack of other distribution media a system evolved, where a scholarly work usually goes through many different hands in a time and money consuming process before it reaches its addressees, mainly members of the same scientific community, who are mostly researchers and writers themselves. As has been pointed out by many authors, these people usually do not publish for economic reasons in the first place, but in order to communicate their ideas and findings to one another (although some other motivations might also be involved). With the Internet becoming broadly used by scholars of all kinds of disciplines and more and more of them directly publishing their works themselves electronically, more reasonable ways of scholarly communication have become visable. As David Green [Green_94] put it: "We might term the Sixteenth and Seventeenth Century was the era of correspondence between great scholars. The Nineteenth Century can be classed as the era of the great societies and the Twentieth as the era of the great journals. The Twenty-First Century will surely become the era of the knowledge web".

Some of the functions of traditional intermediaries, such as publishers, academic societies, press, distributers, bookstores and libraries, can be said to have become superfluous or replacable in an electronic network environment (e.g. typesetting, distribution and marketing), others though are still important (e.g. evaluation and peer review, collection building and long term preservation as well as various forms of author and user guidance), so that we will have to find ways to have them exercised reliably. In the first wave of enthusiasm for the new possibilities of the medium lots of information has been published without much consideration for the search and retrieval side in this worldwide communication space. Additionally to questions of how to structure this information, for academic users it can be a problem that widely recognized systems of evaluation that guarantee a certain selection (like reviewed journals and library collections) have not evolved yet, and it is not certain, whether some of the traditional authorities will take this task in the new medium as well. So here the scholarly communities are facing a lot of challenges, but are also having the chance to create an environment that allows for broader participation and individual solutions.

Thus the aim of this paper is to show a variety of examples of recent scholarly publishing and information selection on the Internet and to identify patterns, trends and possible models (which may also have originated from the business world) that might lead to a desirable scenario of electronic publishing from the perspective of the academic communities. Are there moves to electronic publishing that make thorough use of the potentials of the "knowledge web", do others try to stick to traditional models from the world of paper publishing, or do we find completely different approaches?

The scope of the survey is of course limited: Although the Internet has the potential for a rich use and exchange of academic information in many different formats (like video and audio data), we mainly concentrate on text information here. Neither can we deal with economic and copyright issues in detail.

2. Current patterns of scholarly publishing and communication on the Internet

By "publishing" here we mean all sorts of activities related to making pieces of "intellectual" work available to a broader ("public") audience. Unlike in physical media, in an electronic network context

"publishing" is not just about rather clearly defined "objects" or "items". Creating sense and value out of data is also done by linking, collecting links, annotating and other context creating. In the print world some of this has been the task of libraries, but on the Internet it has become part of the primary publishing itself.

It can also be argued, that we should not view publishing isolated from other scholarly activities. Additionally to private communication and conferences the Internet offers a variety of services suitable for many to many academic communication, e.g. newsgroups, mailinglists, IRC, video conferences etc. An example of a resource collection that recognizes this fact is the "Directory of Scholarly Electronic Conferences" [EConf] (see also the newer versions under [ARLedir]), which lists mailing lists together with e-journals, e-newsletters, etc. In these different communication settings degrees of "formality" and "density" of the information exchanged may vary considerably, so that we might not want to talk of "documents" in every case. Although one can sometimes find the most timely and thought provoking information in more informal communictaion, we would like to concentrate on works, that are intended to reach a broader audience and have a certain degree of permanence.

In the following chapter we try to group academic publishing activities on the Internet according to "players" (who are the creators and publishers) and the kinds of resulting "products" including their arrangement, format, accesibiliy, maintenance and the related communication structures.

2.1 Players and Sites

2.1.1 Individual publication

With the emergence of the Internet for the first time in history it has become possible for individuals to publish works on their own in a medium where they are potentially accessible by millions of people. More and more academics make use of this environment by publishing essays, conference talks, articles pre- and post- print, "e-books", resource guides, observations or collections of links to recommended sites etc. directly at their own homepages. Friends and collegues are notified electronically, and institutions and collegues in the same field just link their resources among one another more or less systematically in a variety of ways. Feedback may be provided via personal mail, discussions on mailing lists or in news groups or in the form of annotations to a web page. In most cases creators of academic content do not publish for economic purposes but rather for sharing ideas and resources, some freelance writers ask for contributions, though. Sharetext projects do exist, but do not seem to be a common scheme.

2.1.2 Geographical or institutional units

Since the University of Minnesota started to establish Gopher as a campus wide information system, many universities, authorities and other institutions have set up servers in order to provide as well local information as pointers to external resources. University or institutional servers today are a main place for publishing works of members of that institution. Within this framework we can find a variety of publishing patterns: In the earlier stages of Internet usage mainly students and jounger faculty of science and engineering departments started publishing information about their institutions and things of personal interest including papers written, course material etc. more or less freely. One can still find this pattern rather often in departments that have traditionally been using few technology. With more

and more "established" people using the net, it has become a common pattern to divide an institution's server into an official area on the one hand, where e.g. publications that have been approved by the center's head are put, and on the other hand a homepage section where everybody publishes and links to places of his interest freely. In recent times we can observe a trend to integrate the information systems of the different departments in one university. Functions of the university's computer center and the library are beginning to merge, and awareness for the need of information management grows. At places with active libraries, like the University of Michigan [UmichLib], Australian National University [ANULib] or Lund University [LundLib], the library may become the central place to help with electronic publishing and management of such collections.

Such a setting encourages faculty to publish "locally". Articles in printed university bulletins used to be hard to retrieve, but they are not any longer in a networked environment. E.g. the University of Virginia Library's Electronic Text Center is running an "Online Scholarship Initiative" [OSI] (also a national pilot project), that allows faculty to publish their works quickly with the help of the center, including guaranteed archiving and linking to other such centers. "Its local mission is to enable UVa faculty to make available on the Internet pre-print copies of articles to be published, post-print copies of articles already published, and occasionally, parallel-print scholarship which serves as an enhanced companion to the print version".

2.1.3 New authorities

In the early days when the Internet used to be an unknown territory for many academics, some sites managed to "claim" certain (publishing) fields, just because they were the first to be there. Nowadays you have to run an especially well designed and maintained, reliable site, in order to get special recognition (in some fields we are still in phase one though). But in both cases places that have proved their ability to adapt to the new technology – mostly this is owed to the personal efforts of individuals –, could become authorities independently of their "status" as publishing places for e.g. paper journals.

An example for a very early site that devoted itself to publishing quality electronic information in the Social Sciences and Asian Studies is the Coombspapers archive of the Australian National University (ANU) [Coombsp]. It was founded at the end of 1991 as an ftp archive in order "to act as an electronic repository of the social science & humanities papers, bibliographies, directories, theses abstracts and other high-grade research material produced (or deposited) at the Research Schools of Social Sciences & Pacific and Asian Studies, Australian National University, Canberra." But as time went by, it attracted other people to deposit their works and consists now (March 1996) of over 2000 ACSII files, amounting to over 70MB of data alltogether. With the event of WWW Coombs archive evolved into the Coombsweb [Coombs] and can still be seen as one of the places with the highest reputation in the field.

In the area of numerical analysis, scientific computing and related fields there is another remarkable example, Netlib [Netlib], "a repository of mathematical software, data, documents, address lists, and other useful items". This digital library, too, started as an ftp archive and is now accessible via WWW. It is maintained by AT&T Bell Laboratories, the University of Tennessee and Oak Ridge National Laboratory with contributors from all over the world. It comes with several search interfaces and can be also searched via the GAMS mathematical software classification system. Netlib started as an e-mail service in the middle of the 1980es, with other protocols coming into use as well in the beginning 1990es. To the day there have been over 10 Million requests to the archive [NetlStat].

One of the great opportunities of a networked environment is the ability to combine publishing with other forms of communication and resource sharing. This has been conceptualized very well by David Green in his idea of the "Special Interest Networks" (SIN) [Green_94]. "A Special Interest Network (SIN) is a set of network sites ("nodes") that collaborate to provide a complete range of information activities on a particular topic. SINS are emerging as an important new paradigm for large scale collaboration on the Internet. Coordination is achieved through logical design, automation, mirroring, standards, and quality control. To be successful, SINs should strive to provide reliable, authorative information services, to encourage participation, and to accommodate growth." Green sees four major functions of a SIN: publishing (text, data sets, imgages, audio, software) at every participating site, virtual library functions (links to topics of interest), on-line services for the community (e.g. analyzing data) and communication through mailing lists, newsgroups, newsletters, conferences etc. In his view, "SINS have the potential to fill both the role of learned societies as authoritative bodies, and of libraries as stable repositories of knowledge and information". They "aim to provide a complete working environment for their members and users", but with decentralized nodes. Other characteristics according to Green include that a SIN meets a special need that is not being met otherwise; it provides coordination and support for the physical network as well as for communication activities; it allows for free access and participation by all its members, which also means that everyone is responsible for editing, formatting, correcting and updating of their own contributions; the members agree on specific data formats, protocols and other standards as well as on a list of usage terms and conditions; and as many processes as possible are automated. It is suggested that one node acts as a kind of secretariat, and that each node takes special responsibilities to coordinate certain projects. Data are published and maintained locally, but every site mirrors or provides links to all important information at the other sites, so that for every access point the database looks essentially the same. For quality control Green suggests several methods including editors applying stamps of approval or ratings to published documents, automated check of incoming data against a database of previous findings etc. Green's own examples of SINs include the Biodiversity Information Network (BIN21) [BIN21] and FireNet [FireNet] (dealing with landscape fires), but similar structures emerge in many other areas as well.

2.2 New roles for traditional intermediaries

Publishers, academic societies and libraries have begun to realize the need to redefine their roles and watch out for new economic models. Committees are set up, strategies announced, alliances are formed with other professions that might possibly loose parts of their old functions in the new publishing game. One example is that publishers and libraries discover issues of common concern and try to work on them together. Publishers look for universities to cooperate in electronic document delivery or digital library pilot projects, etc.

2.2.1 Publishers

Publishers are probably the most endangered species in this game, and at least the bigger ones have become aware of the situation and turned to all sorts of electronic publishing activities. Some have started to distribute supplementary material to paper journals via the Internet and publishing of electronic versions or even electronic only journals.

Elsevier

Elsevier Science [Elsevier] has been experimenting with electronic publishing since the beginning of the 1990s in projects like TULIP, CAPCAS, ELSA or UMDL, mostly in cooperation with US university libraries as pilot users. Building on these experiences Elsevier makes all its 1,100 paper journal titles fully available in electronic form since January 1995 in the "Elsevier Electronic Subscription (EES)" project. For journals that had been included in the TULIP project, backfiles from 1992 onwards are also available. Again the pilot users are big university libraries, many of whom had already participated in TULIP. OCLC is one of the main partners here. EES is totally subscription based, i.e. libraries subscribe to the electronic journals either in addition to or instead of the paper edition.

Journals in this projects come to the project participants on CD-ROM or magnetic tape (with weekly or bi-weekly updates) according to the customer's profile. The packages consist of page images (black/white 300 dpi TIFF), OCR produced unchecked "raw" ASCII files, SGML citation files and a master index containing bibliographic information and pointers connecting the files. Several software packages (e.g. a library's own system or OCLC's Site Search and Guidon client) can be used for access; browsing all pages, zooming and printing as well as searching bibliographic data sets and full-texts (the raw data) are supported. A customer's profile consists of a set of keywords. Each incoming data set (bibliographic and full-text data) is checked against these keywords, and the user is then notified via e-mail about items of potential interest. Pricing depends on the kind of agreement. Flat site licenses are possible as well as pay per download arrangements.

Via WWW Elsevier distributes supplementary electronic information to a number of journals, accessible only for subscribers. Free services on the Internet include browsing of tables of contents of nearly all Elsevier journals, search in bibliographic data and abstracts (but no display of the actual abstract) and delivery of new tables of contents via mailinglist subscription.

Elsevier is also involved in creating a new "Publisher Item Identifier (PII)" standard together with the American Chemical Society, the American Institute of Physics, the American Physical Society and the IEEE. A PII is an extension of an ISBN or ISSN number (added are publication date information and an item number) with the difference that it is assigned to single articles within books or journals by the publisher – thus recognizing the trend to seeing the article as a main unit of publication in an electronic environment.

Springer

Springer's [Springer] focus is also on subscription schemes, with the overall number of electronic journals [Spr_ejourn] being considerably smaller. On the other hand Springer offers a number of electronic-only journals, at least one of them (the Journal of Universal Computer Science, J.UCS) for free. This journal also serves as a joint experiment with the University of Graz for testing their hypertext system Hyper-G [Hyper-G]. In terms of formats, Postscript and pdf dominate. Some of the electronic supplementary material to printed journals at Springer is free.

In 1995 Springer Germany has formed a joint venture with FIZ-Karlsruhe (one of the centers for scientific databases in Germany, STN International member) and a major academic society, the Gesellschaft für Informatik (GI), called the MeDoc project [MeDoc], which is supported by the German Ministry for Science and Education (BMBF) and started with 10 German universities as pilot users. Its aim is to

build an electronic library for researchers and students of informatics. Related is a cooperation between four major German academic societies (in Informatics (GI), Mathematics (DMV), Physics (DPG), and Chemistry (DGCh)) [IuK]. They also seek international cooperation with societies like ACM and IEEE.

AAP

In October 1993 the Professional/Scholarly Publishing Division of the Association of American Publishers (AAP) and the Council on Library Resources (Washington, DC) formed a "Joint Working Group on Professional and Scholarly Information in the Electronic Age", which in 1994 published a report titled "Librarians and publishers in the scholarly information process: transition in the electronic age" [AAPrep]. This report analyzes the roles of all participants in the "information cycle" in detail and stresses the importance of value adding to scholarly publishing by information professionals such as publishers and librarians. Nevertheless they also envision possible scenarios where these intermediaries have lost their current functions almost completely. Some of the very reasonable recommendations include intensive communication between all related parties about new roles and opportunities, shifting focus to content developing, working together with organizations that have an interest in preservation, making out of print archives available, forming task forces together with authors and users etc.

2.2.2 Libraries

As already mentioned above, a lot of university libraries in exercising their function of serving their clients' information needs actively try to incorporate electronic media including networked services. One organizational level higher national library associations are building task forces and try to develop strategies for the electronic age.

ARL

One example of an organization that engages actively in electronic publishing issues is the Association of Research Libraries (ARL) [ARL], which represents over 100 Libraries in the USA and Canada. They distribute the "ARL Directory of E-Journals, Newsletters & Academic Lists" [ARLedir] every year and announce new journals between issues over a mailinglist [NewJour]. A collection of paper and electronic publications is announced via their server [ARLpubl]. One publication of particular interest in this context is a documentation of an e-mail based (mailing lists as well as private) Internet discussion between active ARL members and other scholars and librarians in 1994 concerning the future of academic publishing with the title "Scholarly Journals at the Crossroads: A Subversive Proposal for Electronic Publishing" [Ok-Donn_95]. The original "subversive proposal" by the psychology professor Stevan Harnad suggested that all authors of specialized small area no-market writings ("esoteric publications") should put their works on public ftp servers. The documentation is a good example of how dense and inspiring an e-mail discussion can be – even if put in the shape of a book (with minor editing) and read two years later.

OCLC

Another major player on the library side is the Online Computer Library Center, Inc. (OCLC). OCLC [OCLC] as a "nonprofit computer service and research organization", representing over 22,000 libraries,

mostly in the US, states as one of its important goals to "deliver information electronically to end users, through libraries, from publishers, in ways that enable each to make the transition to a world of information that is both print and electronic" [OCLC_epub].

With Electronic Journals Online [OCLC_EJO], OCLC manages access to electronic versions of journals by several publishers for libraries and their clients. The published list of currently distributed journals comprises nine items, including journals by publishers like the American Institute of Physics (AIP), The Institution of Electrical Engineers (IEE) or Elsevier Science, but one can also find a statement about 60 journals by 11 publishers to be distributed in 1996 [OCLC_epub].

Each electronic subscription requires a contract between publisher, OCLC and subscribing library, where the publisher determines the price. OCLC as the intermediary between publishers and libraries offers to make material available online 48 hours after receiving it from the publisher. The Center handels the technical part as well as all sorts of user queries. Suppliers are free regarding numbering of publications, so there is a certain flexibility as to whether to publish whole issues or single articles.

Subscribers are provided with OCLC's dedicated MS-Windows Client (a Macintosh version is also available) "Guidon", which allows advanced search (boolean, proximity and adjacency, character masking etc. – via a seperate search engine), display of figures, tables and equations, browsing, downloading in ASCII, SGML or Guidon format and printing. In the future more hypertext links to other EJO articles, OCLC's FirstSearch and external databases are going to be added. The underlying database consists of SGML documents, but Guidon as well as the search engine interact with an inverted-file database built from the original one.

Since the beginning of 1995 WWW browsers can also be used as an interface, which means that SGML documents had to be translated into HTML. Roughly speeking, glyphs, figures and tables are converted into gif format images with their own URLs, formula are first extracted form SGML to TeX and then a bitmap is generated. In case a publisher requests a pecial type of characters, OCLC designs corresponding TrueType fonts. For all the different sorts of images a rendering and display engine functions as a bitmap server. It is also planned to add pre-1994 issues as scanned images, viewable with Guidon. (For more information on the software and the translation process cf. [Ingold_94.07] and [Weibel_94]).

2.2.3 Academic Societies

As with libraries, there are huge differences in the attitude towards Internet publishing among scholarly societies in different countries and academic areas. Obviously it is easier in the science and engineering field to adopt new technology for publishing as well. But also among these disciplines there exist traditionally different publishing cultures which seem to determine the attitude to network publishing to some degree (cf. [Hitchc_95.01], [Odlyzko_95], [Ginsp_96.02]). Sometimes it might also be easier for small organizations that have not gained much from paper publications in the past to take innovative steps to network publishing.

ACS

The American Chemical Society (ACS) [ACS] with more than 150,000 members is proud of being "the largest scientific society in the world". It publishes over 20 journals, and its Chemical Abstracts Service (CAS) [CAS] covered about 500.000 articles in 1995. The CAS database is accessible via STN International (operated by CAS, JICST and FIZ-Karlsruhe).

ACS has a long professional experience in electronic publishing. They have taken steps to incorporate electronic material available on the Internet into their databases according to traditional indexing and pricing schemes, but do not seem to experiment with decentralized structures or degrees of formality. One of their main statements with respect to a possible new world of publishing reads as follows: "Without doubt, electronic information networks will play a significant role in the future of science publishing and information sharing. However, while digital files may replace ink on paper, there is no technology that can ever replace the human peer-review and editing process necessary to maintain a reliable science archive for future generations." [ACS_Cyber]. For ACS this clearly means review before publishing and retaining a central database of such quality material.

ACS' recent innovations in electronic publishing include the following: Additionally to the full texts of over 20 ACS journals dating back to 1982, from 1992 on all Chemical Journals of the American Chemical Society are accessible as page images with the software STN Express, which also enables the customer to print out these pages.

Another software package for searching, viewing and printing ACS journals in page image format is called SciFinder. Since 1995 ACS is discussing with Elsevier Science and the Royal Society of Chemistry (RSC) to have some of their journals added for electronic subscription to this system.

Also since 1995 CAS started to abstract and index relevant journals only available electronically on the Internet and to incorparate this information into their frequently updated CAplus database: "Electronic documents, which include journals, conferences where researchers are invited to submit papers, and individually posted papers, will be abstracted and indexed in the same manner that CAS abstracts and indexes printed documents. ... News releases, abstracts or messages placed on electronic bulletin boards or list servers will not be covered." [CAplus]

So-called "Electronic Supporting Information files" for several ACS journals that have previously been distributed on microform are now available online in pdf format for journal subscibers from the ACS server (a large part of the CAS server is now being presented both in HTML and pdf format). Freely available from the ACS-server are full texts of selected new journal articles in HTML.

AMS

The American Mathematical Society has introduced a considerable number of new electronic services in January 1996, all accessible via their Web site "e-MATH" [AMS].

The bigger part of actual research literature is still for fee services. It includes online subscriber access (site license) to the daily updated database MathSci (MathSciNet), which contains bibliographical data from the AMS publications Mathematical Reviews and Current Mathematical Publications from 1940 to the present and full texts from 1980 onwards. AMS' peer-reviewed paper journals are also available to subscribers in electronic form now (for an additional 15% fee; formats are TeX, dvi, PS and pdf), some weeks before the prited issue. Several search functions, hypertext links, access to back numbers and a notification service are provided. All TOCs and abstracts are freely browsable and searchable for non-subscribers as well.

Alltogether ACM offers considerably more free services than the other big societies examined here and its server shows an overall user friendly design, with many useful small services and carefully annotated links to Internet resources of interest to the mathematical community. Free periodical publications in HTML

format include "Mathematical News" and "Mathematical Digest" (smaller newsletter like serials), but also the "Notices of the AMS". All issues of the "Bulletin of the AMS" from 1992 on are available in fulltext, browsable and searchable for free, older issues in TeX and the latest in TeX, dvi, PS and pdf.

Two other electronic only services deserve attention: The "Electronic Research Announcements of the AMS" are AMS' first electronic only journal (since 1995), it is reviewed and freely accessible on the web. Submissions have to be in one of several TeX formats, graphics in Encapsulated Postscript. AMS also runs a free preprint service [AMSPPS], where mathematical preprints available anywhere in the world can be announced. Authors have to provide a set of metadata (including abstract, AMS classification and the URL of the preprint's location – in case the preprint itself is delivered, it has to be in TeX), the resulting database is WAIS searchable and browsable, and there is a mailinglist for notification of new entries. If a preprint has been accepted for publication, the exact bibliographic reference has to be added to the template.

AMS' move to electronic publishing is facilitated by the existence of the long accepted de facto standards of TeX as well as the AMS classification scheme for categorizing.

ACM

The ACM has far reaching plans for electronic publishing. In April 1995 the society's detailed "Electronic Publishing Plan" appeared in the Communications of the ACM and is also accessible from their WWW server [ACM_Epub]. Together with a new "Interim Copyright Policy" [ACM_Copr] this plan is to be revised this year after initial experiments. ACM's long-term goal is to transition all their journals to online distribution, making production and access faster, reducing cost and allowing for more flexible access. Experimental online journals have been set up with free access for a transitory period, but will become subscription based eventually. Freely browsable web pages for every journal then will only contain the tables of contents and general information about the serial.

Archiving in the ACM digital library currently is SGML based. It is hoped for SGML capable browsing and editing tools to become more common some time. In the meantime publications are made available for browsing in HTML, and additionally there will be Postscript files for printing. Publications reaching back to 1990 or earlier are planned to be made available as images only.

Acknowledging current network practice, linking to ACM documents in their central database is highly encouraged, but in case of actual access (=copy), payment will have to be negotiated. ACM concedes that information overload already has lead to an irreversible trend towards the "disintegration of print journals", and access per item – article or even only a small component like a table or figure – is just one logical step further. Under the new copyright policy authors may retain more rights for electronic redistribution of ACM copyrighted material. In case of more than 25% content change a document is no longer considered as a version, but as a new document.

For user needs and economic reasons services and access terms are going to be diversified: Per item payment, subscription and site licenses will all be possible, and interfaces for information retrieval and payment negotiating engines are going to be created. Conference papers and backnumbers will partly be available on CD-ROM. The so-called "Track 1" expert readership paper publications are expected not to be profitable in the long run. On the other hand the new "Track 2" publications for the growing number of non expert readers interested in computing issues are intended to stay in print. In this field ACM plans to cooperate more with other societies like IEEE. Also in this context several new services like

guided access to literature and educational projects (certified knowledge level exams etc.) are planned to be created, and numerous other services are currently being tested.

2.2 Types of publications

In most academic fields paper journals have been the main means of publication during the last decades. For academics who began to use the Internet in the early 1990s, creating "e-journals" was a very natural thing to do. But as the medium (be it a listserve retrieval facility, a Gopher menue item or a hyperlink) allows to point to and retrieve single articles, the notion of the journal as a main entity has become weeker, and unless an editor tries to get together related articles for a special issue, there is not much difference to individual publishing anymore – with review being an important issue though.

We can roughly distinguish between the following common types of arranging electronic publications here:

- independend single publications

- raw data archives

- software archives

- free unrefereed e-jounals

- peer reviewed e-journals (submission, selection and editing process can be exactly like in the paper world), and among those such that are a mere electronically distributed version of a paper journal (for free or with usual subscription and payment, sometimes justifiably more expensive than the print version) and electronic only journals (again these may be free or for fee)

- preprint services (generally free)

- technical reports (some major free services)

- collections of several kinds of items (the form of the single publication may become hardly distiguishable from other items, if journals, archives, pointers to other sources, personal information, search functions etc. are extensively linked.)

Below we give examples of three of the currently most popular patterns of organizing article type electronic publications: (peer reviewed) electronic journals, technical report archives and preprint archives.

2.2.1 Computer Science Technical Reports

NCSTRL

The Networked Computer Science Technical Reports Library [NCSTRL] is an international non-commercial collection of computer science technical reports from about 50 different sites, mainly from US universities that grant PhDs in Computer Science or Engineering, but spreading to other parts of the world and including some industrial and government research laboratories, too.

While the project's predecessors had been supported by grants from the Corporation for National Research Initiatives (CNRI), US National Science Foundation (NSF) and Advanced Research Projects Agency (ARPA), NCSTRL today is an international consortium, the technical support coming from the Cornell Digital Library Research Group.

The aim is to make computer science technical reports freely available to the international research community and to study issues of distributed digital library systems. Participating organizations store their reports locally and provide bibliographic information to the central server, currently at Cornell. They have the choice between running a standard DIENST server or an ftp server (lite site). Generally copyright remains with the authors, and every site has to care for its own legal arrangements. Participants are encouraged to use Postscript or another widely readable format, but they are basically free in their choice. The user thus gets page images, PS files, HTML and other formats. The reports are searchable for keywords in the abstracts, author, title, document identifier and institutions via a WWW browser.

UCSTRI

Marc VanHeyningen of Indiana University with his Unified Computer Technical Report Index [UCSTRI] takes a different approach to search and retrieval of distributed technical reports. In this system an index builder polles a list of ftp sites every week to create a master index. A search engine then checks this file for keywords received through a WWW search form and returnes URLs together with bibliographical information or abstracts to the user. No cooperation from the site of the ftp archive maintainers with respect to software or metadata format is required, because the indexer extracts this information according to supposed common naming habits. So generally the returned items do not appear in a nicely formatted form (though there is a useful feature of arranging the hits according to date), but the potential number of resources is much bigger than in an archive maintainer supported project. Currently the indexer lists nearly 15.000 items from 200 sites.

2.2.2 Preprint Services

Physics seems to be the discipline with some of the oldest and most widely used electronic preprint services, the most famous being "xxx.lanl.gov". The "xxx.lanl.gov e-Print archive" at Los Alamos National Laboratory [LANL_eprint], an automated system for announcing, archiving and retrieving electronic preprints, was initiated by Paul Ginsparg in 1991 and today (February 1996) processes more than 70,000 electronic transactions per day. Although it first started as a high energy physics service, the archive has extended to other areas of physics, to several mathematical disciplines and also less related scientific fields and consists of more than 20 different subject archives today.

Preprints are submitted via e-mail with some sort of TeX resp. LaTeX being the default format (PS and pdf are accepted under certain conditions; graphics in (Encapsulated) Postscript), then TeXed automatically and also converted into Postscript. Subscribers of an archive's notification service receive an e-mail with author, title and abstract information of every new paper on the day of submission. Withdrawel and updating of preprints also is handled via specific commands. Users can search the archive and retrieve preprints in any one of the formats TeX source, hyperdvi, gzipped hyperPostScript or pdf (some archives do not carry all of them).

According to Odlyzko [Odlyzko_95] it took only one year for the scientific community of high energy theoretical physics to almost completely switch to Ginsparg's system as the primary means of information dissemination. Ginsparg [Ginsp_96.02] adds that in some areas of physics the archives "have already supplanted traditional research journals as conveyers of both topical and archival research information."

In accordance with the rich preprint culture in physics in general we can find other big preprint services, e.g. at CERN [CERN_prep] and also efforts to provide a unified interface for all of the most important sites [ICTP].

In mathematics, too, preprint services are very popular, and efforts are being made to provide unified services here as well. The AMS e.g. offers registration and search facilities for mathematical preprints from all over the world [AMSPPS].

It does not surprise not to find many such sites in the humanities and social sciences, where preprints have not played a big role in print publishing either. Nevertheless services like the "International Philosophical Preprint Exchange" [IPPE] at Chiba University do exist.

2.2.3 Scholarly electronic journals

Before peer reviewed scholarly electronic journals became popular on the Internet, there already existed a large number of so-called "e-zines", electronic magazines with content ranging from academic material to pop culture. Many directories of electronic serials, like the ARL list of "journals and zines" [ARL_jz], still mix all kinds of content, but as the number of publications grew, categorizing according to subject areas has become more common. Today we can find a number or collections of pointers to academic e-journals.

The Committee on Institutional Cooperation's [CIC] Electronic Journals Collection [CIC_EJC] used to list about 800 electronic journals in a less systematic way, but is currently being restructured. It ultimately "aims to be an authoritative source of electronic research and academic serial publications – incorporating all freely distributed scholarly electronic journals available online". A master database of bibliographic records, browsable (WWW) and searchable (WAIS) is being created. It is planned to incorporate journals, that are only licensed to CIC member universities into the collection later.

The Scholarly Societies Project of the University of Waterloo Electronic Library maintains a list of about 130 Scholarly Society Serial Publications [SSP_Arch], many of them newsletters, but also including reviewed journals.

There also exists a meta index to academic (and other) e-journal sites at The University of British Columbia Library [Ac_Ejourn], which makes for a good starting point.

It would require a seperate study to evaluate content, structure, innovative ideas etc. in all the publications available today, but fortunately we can refer to a very useful survey that has been conducted between September and October 1995 as part of the UK Open Journals Framework project, which examined the state of (English language) online journals in the science, technology and medicine fields [Hitchc_95.01]. In the STM field alone over 100 online full-text, peer-reviewed journals were found, about half of them electronic editions of paper journals. Over a third of them were only accessible for fee (or are scheduled to be). Half of the journals first appeared in 1995, and hardly any dated back to pre-1993. (According to Stevan Harnad, "Psycoloquy" [Psych], which he started in 1990, is the oldest peer-reviewed scholarly journal on the Internet.) A general finding was that only a few of the electronic journals made thorough use of the many new possibilities of the medium, like hypertext markup, links to external sources, use of video and audio data, electronic notification services, electronic delivery earlier than the print version, new journal structures etc. Older journals used to be completely in ASCII format, and some people argue that ASCII is still important for archiving (the same argument is given for bitmaps) and access

from less developed regions. HTML is very common in more general publications, that do not need many equations or tables. In addition to Postscript recently pdf format has also become popular, whereas only one of the examined journals used page images. Notable differences exist between the various disciplines: Mathematics journals had the biggest share in this survey, most of the journals being electronic-only. TeX, dvi and Postcript predominate – as in paper publishing before. For physics the point is made, that the Los Alamos e-print archive – building on an existing preprint culture –, is so dominant, that only few electronic-only journals emerged. In addition to the formats used in mathematical publications, HTML and pdf were also common. In biology and medicine the need for graphical representation is reflected in prevalent use of HTML and pdf formats. Major journals in medicine still seem to be reluctant to go online. Computer science is said not to have build highly organized publishing structures yet, but to rely more on a number of ftp archives and a conference culture, so few electronic journals resulted so far. As for funding electronic journals the survey does not see a general trend yet. Big publishers are supposed to stick to the subscription model with all sorts of pricing being experimented with. Pay per view is considered to be difficult, author page charges do not seem to be very popular, but totally free access is also judged to be dangerous for a healthy long life of a journal.

2.2.4 Some general trends in systems and formats

In the examples above we have seen a number of approaches to free but organized publishing. Whereas at Los Alamos a central archive of e-prints is maintained, the CSTR services as well as AMS' preprint service leave the actual papers decentralized and provide only a central database of pointers. There are advantages and disadvantages with both models. Most of the examined peer-reviewed journals still follow traditional publishing patterns, but since WWW browsers have become the standard interface, more of them also try to experiment with internal and external linking, timely instead of issue wise publishing, including data other than text etc.

The question which format to use for presentation is highly related to the specific needs in the respective disciplines. Whereas ASCII and HTML often is sufficient for social sciences and humanities (with gif and Postscript being used for graphics), mathematicalized disciplines need to present equations etc., so TeX is still considered the most suitable format. In the submission guidelines for the Los Alamos archive it is explicitly stated that TeX is prefered, because it has high capabilities for transporting structural information, it is fully searchable in the (ASCII) source, easily to distribute via e-mail and translatable to hypertext and graphical formats. In areas outside the sciences it is not widely accepted though.

Recently Adobe's "portable document format (pdf)" (cf. [Acrofaq]) has gained a lot of popularity, even in disciplines where most people are capable of handling TeX. Pdf has been created on the basis of Postscript, but is designed to be platform independend. A pdf file includes e.g. page descriptions (concerning the arrangement of text, graphics, images) and font metrics; font substitution is possible without reformatting. Pdf documents are structured into "objects", can include hyperlinks, and allow for annotations. They are usually smaller than Postscript files, one reason being image compression. Printing is possible from any Postscript printer, but only via conversion into PS format. Using hypertex and extended dvips TeX can be translated into pdf, probably one reason why mathematicians have come to use it, too. Adobe freely distributes a software called "Acrobat Reader" for viewing pdf files. This reader has a "find" function for searching strings in a pdf file, a major improvement compared to Postscript, but more sophisticated searches can only be performed with a commercial search engine

called "Verity". With this product Adobe explicitly aimes at Internet publishing applications. Pdf is portable, graphics, URLs and security information can be incorporated in a document, and Acrobat works together with common WWW browsers. While earlier versions of the reader were started from within Mosaic of Netscape, version 2.1 is intended to be able to follow links itself. All single-byte fonts are expected to work correctly with Acrobat 2.0, but double-byte font support is planned only for the Acrobat 3.0 products, scheduled for 1996.

3. Current patterns of information choice

Busy people often complain that they just do not have the time to spend several hours a day "browsing" the Internet for the professional information they need. While – like in the world of print – it depends a lot on what you are looking for (examples might include some quick bibliographical reference, experts working in your field of interest, general or specialized, old or very new material, text or other data or material in a special language), it is true that systems and customs in Internet publishing are not well enough established yet for letting people develop a feeling for what is available and where to find it. In reality – with paper as well as electronic material – people probably apply all sorts of mixed search strategies, the simplest way being to ask someone you know who knows ... . Browsing "geographically", i.e. going to a place where you suppose appropriate information to be, is also an important pattern, that will not vanish because of improved direct search of keywords in full texts.

Already now several kinds of information are much easier to find on the Internet than in print (or for that matter for the first time remotely accessible at all), e.g. personal information including references to all types of works by a certain author (once you know her homepage), all kinds of less formal but important information, library catalogs throughout the world etc. With the growing amount of networked information gradually search services also have been developed, first on the basis of the different protocols (X500 lookup or Whois for mail addresses, Archie for ftp, Veronica for Gopher, searchable WAIS databases, and recently all kinds of web crawling engines for the WWW). Nowadays many combined search facilities with a WWW interface have become popular, and many of them try to integrate the two basic approaches of providing subject catalogs and full text searching. Still the usefulness of the returned material varies greatly from tool to tool. In many areas the kinds of information academics look for (works in specialized areas, which requires detailed subject classification, and works of a certain quality, which requires evaluation standards) are not sufficiently searchable yet.

A serious problem with most of the currently used Internet protocols is that they have not been developed for the management of global scale digital libraries. Nor do most authors even use the given markup possibilities thoroughly and correctly, which also limits the usefulness of all currently available search tools considerably.

On the other hand, if used efficiently, the electronic network environment opens up many new search possibilities in addition to traditional search patterns (subject or author search, browsing potentially relevant journals or querying an authority etc.): besides automated keyword and context search in huge data masses there are also possible ways of more personalized search like individual and group filtering, but most of them are not widely used yet.

In the following we would like to introduce the currently most common tools for support of Internet search along with pointing out some of the problems and advantages related to each approach: subject catalog building and searching, using several kinds of search engines, and searching information that

has been evaluated in some way (actually search for a special content and search for a certain quality technically do not necessarily have to be different things – this depends e.g. on the kind of markup –, but lacking examples of successfully combining these categories, we will deal with them seperately.)

3.1 Subject catalogs and guides, browsing in "virtual libraries"

Subject catalogs on the Internet vary in terms of the range they try to cover (from attempts to classify potentially any existing information, via catalogs for a geographical or organizational unit (country, university) to guides in a certain academic area) and the degree of "involvement" with the resources they present.

The W3 Consortium's "WWW Virtual Library" project [W3VirtLib] dates back to the year 1991 and is an example for an attempt to coordinate "virtual libraries" for a wide area of subjects by assigning the management of a special field to distributed managers. The only requirements for participating web masters are sticking to a uniform look of all VL pages and refraining from advertising. In one variation this catalog is arranged according to Library of Congress categories. [W3VL_LOC] Examples of well managed partial libraries under this umbrella are the Asian Studies [ANU_AsStud] and Information Quality Libraries [ANU_Qual] at Australian National Universiy (ANU).

In a similar approach the "Clearinghouse for Subject-Oriented Internet Resource Guides" at the University of Michigan [UmichClear] provides a central place for subject guides to an area of the compiler's choice (subject and range can be determined freely as long as it meets formal requirements). Allthough as their number grew, the guides were grouped according to some top categories as well, no certain overall classification scheme is imposed.

Different to the two examples above, Yahoo! [Yahoo], started in 1994, does not leave the classification in subgroups to individual (volunteer) maintainers, but tries to maintain a consistent subject hierarchy for any information on the WWW, employing about 20 professional classifiers, who continuously discuss categorization among each others. [Steinberg_96.05]

There are many more examples for useful Subject indices in special areas, e.g. the Engineering Electronic Library, Sweden (EELS) [EELS]. This is a joint effort of six Swedish University of Technology Libraries to collect evaluated pointers (assessed are factors like "accessibility, maintenance, documentation and reliability of producer organization") to Internet resources for Engineers. An editorial team of about 10 people active in the field is in charge of the collection – each in certain subareas – and they classify the approved resources according to a scheme by the US based Engineering Information Inc. Page titles and descriptions are WAIS searchable.

One of the European Union funded projects within the Telematics Applications Programme, Telematics For Research is called "DESIRE" (Development of a European Service for Information on Research and Education) [DESIRE]. This 3 Million ECU project has started earlier this year and is scheduled initially for 27 months. DESIRE aims at creating a consistent European WWW index for browsing research and education related information available anywhere in Europe as well as at providing so-called "subject-based information gateways" (SBIGs) to independently managed quality controlled specialized collections. A number of tools for creating and maintaining meta-data (URC), for information discovery systems, facilities to manage closed user groups etc. are also to be created. Another task of DESIRE is to focus European discussions concerning standardization. Being a WWW based project, it is planned

to adhere to existing and emerging standards set by the W3 Consortium and the IETF, and European requirements (like multi language capabilities) are to be addressed to these committees.

3.2 Engine supported "direct" search

A natural reaction to the rather unstructured publishing habits on the Internet was the creation of search engines that do not rely on much structured information, but have the ability to search large data masses for certain keywords. Recently the number and character of different search engines itself has become so diverse that it is hard to get an overall view of them. This situation in turn resulted in numerous comparative studies of different range and quality. Here the works by [Koch_96.03], [Koch_Search], [Steinberg_96.05], [Zorn_96.05] and [Courtois_96.05] proved to be very helpful.

The many engines differ in a number of ways. One is the "geographical" range of the data searched, which reaches from information stored in a certain file system on a single server, via data belonging to a common domain to virtually all URLs throughout the World Wide Web. Recently services of combined search of Internet resources and traditional databases, usually not reachable by freely crawling engines, have also appeared on the scene. An example is IBM's "infoMarket". [infoMarket]

Another important difference concernes the types of data searched. Some engines search for human assigned categories, keywords, titles/headders or abstracts (supplied either by an index editor/compiler or by content creators who register their site and deliver those metadata themselves). Basically this is an extension of traditional library search. Among these template using engines are Yahoo!'s search or CUI's W3catalog [CUIW3cat]. Advantages are a certain degree of reliability for users, if there exists an overall classification scheme – like at Yahoo! – (orientation within consistent categories) and ease of search for data that are difficult to search directly, like image or audio data, compressed or encoded data. Disadvantages are a possible ontological and categorizing bias and eventually consistency problems because of the information masses (a single site cannot cover everything, and different people categorize differently).

Full text keyword search engines like Alta Vista [AltaVista], Lycos [Lycos] or WebCrawler [WebCrawl] have the advantage that what they return is based on real occurence and not classifier and scheme dependend. Disadvantageous on the other hand is the missing context, the fact that "real aboutness" and relevance of keywords cannot be grasped this way, synonymes are not recognized etc. Some of these engines, like Alta Vista, offer advanced search functions like boolean and proximity search, truncation and ranking of returned documents.

Additional computer generated information on a fulltext indexing basis like statistical context analysis, elimination of homonyms (also based on statistical relations) etc. can help to focus the search further. Examples are excite by Architext [excite] or Magellan [Magellan]. These tools also use a (probably human made) synonym thesaurus. Here we can avoid the classification bias, and the system is adaptable to newly emerging contexts, but still statistical relatedness does not guarantee for semantic relatedness.

So full text search and statistical cooccurency analysis surely is a major improvement compared to bibliographical search, but when it comes to "real" context, especially in the humanities and even more so in literature, for human like understanding there probably has to be some human input, which means the concepts and categories depend on the participating individuals. One attempt to create a huge context "knowing" database as the basis for automatic classification is Oracle's ConText [Steinberg_96.05].

Classification related work has for a long time largely been done by a limited number of librarians (though in some fields the creation of classification schemes has been a joint effort of at least a number of experts). But with the "volume of knowledge" becoming bigger and bigger, the work of classification should be shared between more experts, each contributing in her field of knowledge. The creation of context databases for automatic search support might be another way of looking at classification. Here the scientific communities resp. scholarly societies/digital libraries in certain fields could surely contribute a lot from their experience. Like with machine translation there may be a lot of machine support possible, but eventually sense is created by humans for humans.

Collected and combined search tools have become very common recently. With more and more commercial search tools competing, every site tries to provide many different approaches at once (subject catalogs, pointers to other guides, robots, hotlists) and to look as comprehensive as possible. Whereas in earlier Internet days "Meta Indices" like Mosaic's [MosaicMeta] pointed mainly to catalogs, newer collections of starting points like Netscape's NetSearch [NetSearch] focus almost entirely on search engines. "Meta" now means creating combined search engines that query several other engines and indices at a time (e.g.MetaCrawler [MetaCrawl]).

While the growing number of universal crawlers often imposes quite a burden on sites that are "visited" by them, more intelligent distributed search tools are also being created. One of the more intelligent tools in this respect is Harvest with its flexible and distributed structure. Harvest was originally designed and built by the Internet Research Task Force Research Group on Resource Discovery (IRTF-RD) and can be accessed from the University of Colorado [Harvest]. It is an "integrated set of tools to gather, extract, organize, search, cache, and replicate relevant information across the Internet." [Harvestfaq]. Harvest consists of so-called gatherers (running locally at a site to gather information regularely and then provide updates, which prevents sites from becoming overloaded from indexer robots), brokers (which query multiple gatherers; there are special brokers for different subject areas), index and search subsystems, replicators and object caches. Harvest allows you to use several search facilities of your choice like WAIS, Glimpse, and Nebula. Glimpse is the default, because it supports search functions like and/or queries, approximate searches, regular expression search, case (in)sensitive search, matching of parts of words, whole words, or multiple word phrases, variable granularity result sets etc. Information from all protocols can be indexed on a full text basis, and Harvest runs different "extractors" or "summarizers" that extract information from various data types (e.g. PS or TeX files, tar files; there currently is also an experiment to extract pdf files via conversion into PS).

3.4 Searching for evaluated information

As network users in our information choice we can consider evaluations by an "authority", by an anonymous group of similar minded people or by ourselves. (For a general collection of quality evaluation approaches see [QualGuidel].)

3.4.1 Authority rated information

Most existing schemes of "quality metadata" so far use the "authority" approach (evaluation by experts or institutions). Ratings can be applied "internally" or "externally" and different technical approaches are possible.

Very common are "semi-external" ratings, by which the editors of a collection of pointers (not of the material itself) evaluate each pointer, and assign to each site or item "seals of approval" (a concept initially developed in the Interpedia project [Interped]) or a certain rank. A corresponding logo can then be "tacked" to the site itself. The Clearinghouse at the University of Michigan has an editorial team that evaluates the guides listed at their site at least once a year according to the criteria "level of resource description, level of resource evaluation, guide design, guide organizational schemes, guide meta-information" (each feature seperately on a five point scale) and chooses a "guide of the month" [Umich_rate]. At present, badly rated sites are not excluded from the collection. McKinley's staff evaluates sites for their Magellan directory [Magellan] according to the categories "depth", "ease of exploration" and "net appeal" on a scale of 1-30 points or 1-4 stars [MagellFAQ]. The evaluated site is then allowed to show the respective logo. Point [Point] Web Reviews, connected to Lycos and A2Z, presents their "top 5% of the web" in a number of categories, rated on a 50 point scale in five categories according to "content" (broad, deep, thorough, accurate, up to date...), "presentation" and "experience" (fun, worth the time). The annual "Best of the Web" award [BestofWeb] and GNN's Best of the Net Awards [GNNBest] are further examples.

Compared to these patterns, the "PICS filtering scheme" [PICS], a joint effort by the W3 Consortium and representatives of over 20 companies since August 1995, provides a basis for entirely "external" ratings, where labels can be distributed independently of sites. Originally created for parents' "flexible blocking" of offensive Internet content according to categories of their choice, the "Platform for Internet Content Selection" is an attempt to establish conventions for label formats and their distribution methods. Since PICS allows for all kinds of labelling services and filtering software to be used together, it gives individuals an opportunity to choose one or more rating schemes as well as software by different makers. PICS labels can be distributed in several ways, e.g. by authors as embedded labels in HTML documents, by a label server connected to an http server (publisher's labels) or by a third party "label bureau" that can be asked for evaluations (such label collection could also be distributed through other media like CD-ROM). Additionally to PICS' primary aim a lot of applications can be imagined (cf. [[PICSoutl]). Although labels at the moment cannot contain arbitrary text, they could instead point to a URL and thus transport all kinds of information. It has been suggested to distribute subject category or quality information, information on copyright, distribution rights or request for payment via such labels, to use them for rating single articles in e-journals or also for collaborative filtering (see next section).

3.4.2 Collaborative filtering

A number of collaborative filtering systems are currently developed. Some are based on statistical "agreement" about quality aspects (and thus use simple scales for rating), others provide the possibility of extensive annotations to documents.

Statistical group filtering is based on the concept that some people tend to rate the same information in a similar way, but that these people do not necessarily know each other. With the help of statistical analysis of common ratings predictions concerning the potential usefulness of particular objects can be made. (This method corresponds to the cooccurence concept in content search mentioned above).

One example for anonymous statistical group rating is the project GroupLens [GroupL]. GroupLens is a system for collaborative rating of Usenet news articles. Every participant gives a brief rating information (a number between 1 and 5) about an article he/she has just read and receives, on the basis of automatic

comparison of one's own ratings with those of many other people, a prediction for unread articles, drawn from the ratings of people one usually agrees with. In GroupLens there is a mechanism to keep individuals anonymous, but the same system could be used to share ratings with specific people.

For more elaborate systems of shared comments cf. [Roesch_95]. Within the MIT Agents Group there are also being developed a number of "matchmaking" agents [MIT_Agents].

3.4.3 Personal Agents

"Butler type" instructable agents that support entirely individual search by using a profile, previous ratings etc. are just being developed, but will probably play an important role in future search patterns (cf. the various projects at MIT Media Lab, Autonomous Agents Group [MIT_Agents]).

4. Conclusions

Some desirable ingredients of a future world of academic publishing are already being tested in many corners of the Internet, in a number of cases, as in the Los Alamos e-print example, rather successfully, i.e. well accepted by the respective scientific community. Nevertheless it is not easy to predict, whether most other disciplines will follow in a similar direction and what kinds of publically supported new systems might emerge out of this. It is up to the many academic communities to create and experiment with new forms of publishing suitable to their respective needs, that might evolve into models convincing enough to eventually gain enough support from their funding organizations. The development of Internet technology in general is a good example for what the joint effort of highly motivated geographically distributed individuals (and supportive organizations) can achieve compared to official international standardization committees. The outcome is likely not to be the most sophisticated, but rather the most practicable solution.

From the perspective of people primarily interested in sharing ideas, we hope for open systems to evolve that allow for many people from different areas and with different motivations to contribute, but at the same time provide powerful tools to choose the information suitable for everyone's individual purposes as well as flexible rewarding schemes. The Internet enables us to work geographically decentralized and still cooperate worldwide with few institutional overhead. Nevertheless institutions play an important role in providing the various levels of infrastructure we rely on for our activities. Also an especially important task now is the creation of institutional frameworks for new public functions like maintaining reliable electronic libraries and archives.

If we look at the factor openness, besides entirely independend publishing, the various examples of preprint services, technical reports and software archives as well as some electronic journals seem to be the most advanced solutions. In most cases these archives are at least searchable for author assigned metadata, and it is encouraged to use widely usable formats. They are very popular in spite of the fact that no review takes place before publication. This may depend partly on the field. In some areas timely information is so important, that less carefully prepared drafts will be accepted. In areas where more general or less time sensitive information is needed, people might rather turn to services that only announce or archive certain quality selected material. But even if due to the varying publishing and communication cultures notification services may take different shapes in different disciplines, the overall number of drafts, preprints, requests for comments etc. published in an early stage of a work,

prior to review can be expected to grow. It has been argued that a built-in feature of archiving registered preprints without a possibility of withdrawel might be a good incentive for quality (self) control before submission [Odlyzko_95], but on the other hand the possibility to apply corrections and updates has considerable advantages, if version numbers and "diff" files are supplied.

In some academic disciplines (e. g. such that cannot build on an existing preprint culture or where distribution of inaccurate information might have serious consequences), at least in the nearer future probably reviewed and edited electronic journals, i.e. selection by authorities, will stay popular. Because this is a form most people are acquainted with, it is also suitable for cautious moves into network publishing. If more scholars of a certain reputation decide to edit an electronic journal, especially jounger academics will not have to worry too much that publishing electronically will hinder their carreers. Currently most electronic-only journals are run by individuals or academic institutions rather than commercial publishers. A general problem with all the independend archives is that each may have its own preservation policy, so there is mostly no insurance that the material will last over time.

So far commercial publishers and big academic societies are still in a good position to sell electronic journals, partly because they have a huge amount of (already published) data at their disposal that can be turned into new electronic products. In this centralized model handelling of formats, internal linking and charging flat rates for access to a certain database is easy. While for-profit publishers tend to stick to a subscription model to ensure income, societies like ACM take a step further to pay per access schemes and encourage extensive linking with external resources. But even if we envision huge parts of academics to move into free electronic publishing so that the amount of available information will soon have surpassed the commercially published, people probably still will be willing to pay for special quality products. So commercial publishers could concentrate on special collections, "brand" development or archive maintanance. Also we can expect all sorts of smaller intermediary and value adding services evolving, for e.g. helping with document creation, abstracting, markup and layout on the one side, helping with retrieval on the other.

Supposed some time there really will be most academic information available for free (and methods of direct payment facilitated to those authors who do want to charge a certain amount), what models might evolve on a larger scale? The preprint notification and distribution scheme works fairly well for relatively narrow academic areas, where people know exactly which notification service to subscribe to and the type of data exchanged is confined to a more or less unified format. If we look at information and communication infrastructures in larger areas, where information of different kinds is involved, approaches resembling Green's concept of Special Interest Networks seem to be more fruitful. A lot of the tasks arising here could be taken care of by information specialists from the merging organizational forms of academic societies and (digital) libraries. In order to build user friendly collections, interdisciplinary cooperation between experts in the different fields (who are authors and users at the same time), general information specialists and technichal experts is desirable. With authors and search specialists working closer together, more effective and personalized search methods might be developed.

Still there remain a lot of open questions, if we envision distributed sites to store and manage the biggest part of available electronic information. One is the question, whether there should still be central "national" archives or whether instead such alliances could be entrusted with public tasks like managing public libraries/archives. Since discipline networks are getting increasingly international in character, this is also connected to the problem of international interdependence and to the question what kind of a policy every country should take regarding material in its primary language. If we still

need central archives, what would be their role? Should they store what the experts of the respective academic communities consider worth preserving, should they make a selection of their own or should they indiscriminately store any information electronically available (because here space is not a crucial issue any more)? Although preserving virtually everything might not be a big problem technically, keeping such archives usable is a different question. Full text search over Terabytes of data probably will not bring satisfactory results in many cases, so additional structure, at least in the form of "recommendations", will be needed.

If we are talking about more or less "national" academic societies/digital libraries, it might not be too difficult to give the responsibility of managing an "official" collection in a specific field to such a networked society (or a smaller chapter of a society, maybe in cooperation with some other institutions) with public funding. It then would have to guarantee access to material that does not find enough organizational support otherwise and to maintaining general archives. A multiple step screening process for several degrees or categories of "preservation worthiness" might be established. Such a multi step process could have the advantage of serving different usage needs and different time perspectives: Material that is considered to be of a special quality, could be marked up and classified more carefully and thus kept better accessible than the rest. Also what is needed "now" might not be equally useful for generations to come. So evaluation processes for immediate usage should take fewer time than long time usage evaluations. In the latter case, it is more important to invest some more time and thoughts in order to choose the "right" items. Probably it would also be a good idea to have an interdisciplinary committee coordinate the various long term preservation activities, so that a possible bias may be balanced. Maybe a society could take the duty to keep "middle term" archives, and within this period the overall preservation specialists decide about long term preservation (such a procedure might have technical benefits as well).

The overall situation now looks like authorities will still need some time to develop and decide upon general preservation policies of networked information comparable to national library collection guidelines.

Network publishing will become more attractive with better technology available for a couple of tasks: Further development in the direction of formats like pdf, that combine hyperlinking with advanced possibilities of displaying graphs, tables and equations, while keeping text searchable and being reasonably easy to use by authors, is likely to occur. Flexible and secure tools for charging will be helpful in some areas. Search methods can also be expected to be improved. While the effectiveness of new selection methods will partly depend on the willingness of authors or third parties to provide metadata in certain formats that are usable by search engines, or of users to cooperate in collaborative filtering, progress will probably also come from distributed gatherers and advancements in sheer data amount handling capacity.

URLs

[Ac_Ejourn] Ejournal SiteGuide: Academic

`http://unixg.ubc.ca:7001/0/providers/hss/zjj/ejacd.html`

[Acrofaq] Adobe Acrobat Frequently Asked Questions

`http://www.adobe.com/acrobat/acrofaq.html`

[ACS] American Chemical Society

`http://www.acs.org/`

[ACS_Cyber] The Race for Cyber-Space

`http://pubs.acs.org/journals/wspp/race.html`

[AltaVista] Alta Vista

`http://altavista.digital.com/`

[AMS] American Mathematical Society

`http://www.ams.org/`

[AMSPPS] AMS Preprint Server

`http://www.ams.org/preprints/`

[ANU_AsStud] The World-Wide Web Virtual Library Asian Studies

`http://coombs.anu.edu.au/WWWVL-AsianStudies.html`

[ANULib] The Electronic Library and Information Service at ANU

`http://info.anu.edu.au/elisa.html`

[ANU_Qual] The World-Wide Web Virtual Library Information Quality

`http://coombs.anu.edu.au/WWWVL-InfoQuality.html`

[ARL] Association of Research Libraries

`http://arl.cni.org/`

[ARLedir] ARL Directory of E-Journals, Newsletters & Academic Lists

`gopher://arl.cni.org/11/scomm/edir`

[ARL_jz] Electronic Journals and 'Zines

`gopher://arl.cni.org:70/00/scomm/edir/edir95/jz`

[ARLpubl] ARL Publications Catalog

`http://arl.cni.org/pubscat/publist.html`

[BestofWeb] Best of the Web

http://wings.buffalo.edu/contest/awards/

[BIN21] Biodiversity Information Network

http://life.csu.edu.au/bin21/bin21.html

[CAplus] CAS announces first comprehensive effort to offer scientists access to electronic-only documents

http://info.cas.org/New1/epub.html

[CAS] Chemical Abstracts Service

http://info.cas.org/

[CERN_prep] CERN Preprint Server

http://preprints.cern.ch/

[CIC] Committee on Institutional Cooperation

http://cedar.cic.net/cic/

[CIC_EJC] The CIC Electronic Journals Collection

http://ejournals.cic.net/

[Coombs] Coombsweb - ANU Social Science Server

http://coombs.anu.edu.au/

[Coombsp] Coombspapers

http://coombs.anu.edu.au/CoombswebPages/Coombspapers.html

[CUIW3cat] CUI (Centre Universitaire d'Informatique, University of Geneva) W3 Catalog

http://cuiwww.unige.ch/w3catalog

[DESIRE] Development of a European Service for Information on Research and Education

http://www.nic.surfnet.nl/surfnet/projects/desire/

[EConf] 8th Revision Directory of Scholarly Electronic Conferences March 1994

http://www.lib.ncsu.edu/reference-acadlists.html

[EELS] Engineering Electronic Library, Sweden

http://www.ub2.lu.se/eel/

[Elsevier] Elsevier

`http://www.elsevier.nl/, http://www.elsevier.com/`

[excite] excite

`http://www.excite.com/`

[FireNet] The International Fire Information Network

`http://www.csu.edu.au/firenet/`

[GNNBest] GNN Best of the Net Awards

`http://nearnet.gnn.com/gnn/wic/botn/index.html`

[GroupL] GroupLens

`http://www.cs.umn.edu/Research/GroupLens/`

[Harvest] The Harvest Information Discovery and Access System

`http://harvest.cs.colorado.edu/`

[Harvestfaq] Frequently Asked Questions (and Answers) about Harvest

`http://newbruno.cs.colorado.edu/harvest/FAQ.html`

[Hyper-G] Institute for Information Processing and Computer Supported New Media (Hyper-G server)

`http://hyperg.iicm.tu-graz.ac.at/`

[ICTP] International Centre for Theoretical Physics, Trieste, Italy, One-Shot World-Wide Preprints Search

`http://www.ictp.trieste.it/indexes/preprints.html`

[IPPE] International Philosophical Preprint Exchange

`http://phil-preprints.L.chiba-u.ac.jp/IPPE.html`

[IuK] Gemeinsame Initiative der Fachgesellschaften zur elektronischen Information und Kommunikation IuK (in German)

`http://alice.physik.uni-oldenburg.de/IuK/`

[infoMarket] IBM infoMarket

`http://www.infomkt.ibm.com/`

[Interped] Interpedia Project

`news: comp.infosystems.interpedia`

[Koch_Search] Literature about search services

`http://www.ub2.lu.se/desire/radar/lit-about-search-services.html`

[LANL_eprint] xxx.lanl.gov e-Print archive

`http://xxx.lanl.gov/`

[LundLib] Lund University Electronic Library

`http://www.ub2.lu.se/`

[Lycos] Lycos

`http://www.lycos.com/`

[Magellan] Magellan

`http://www.mackinley.com`

[MagellFAQ] Magellan Frequently Asked Questions

`http://www.mckinley.com/feature.cgi?faq_bd`

[MetaCrawl] MetaCrawler

`http://metacrawler.cs.washington.edu/`

[MeDoc] The Electronic Computer Science Library

`http://medoc.informatik.tu-muenchen.de`

[MIT_Agents] MIT Media Laboratory Autonomous Agents Group

`http://agents.www.media.mit.edu/groups/agents/research.html`

[MosaicMeta] NCSA Mosaic's Internet Resources Meta-Index

`http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/MetaIndex.html`

[NCSTRL] Networked Computer Science Technical Reports Library

`http://www.ncstrl.org/`

[Netlib] Netlib Repository

http://www.netlib.org/

[NetlStat] Netlib Statistics at UTK/ORNL

http://www.netlib.org/utk/misc/counts.html

[NetSearch] Netscape Net Search

http://home.netscape.com/home/internet-search.html

[NewJour] NewJour Announcement List

gopher://arl.cni.org:70/00/scomm/edir/newjour/

[OCLC] OCLC Online Computer Library Center, Inc.

http://www.oclc.org/

[OCLC_EJO] Electronic Journals Online

http://www.oclc.org/oclc/menu/ejo.htm

[OCLC_epub] Electronic Publishing

http://www.oclc.org/oclc/ar95/epub.htm

[OSI] Online Scholarship Initiative, University of Virginia Library's Electronic Text Center

http://etext.lib.virginia.edu/osi/

[PICS] Platform for Internet Content Selection

http://www.w3.org/pub/WWW/PICS/

[PICSoutl] PICS: Internet Access Controls Without Censorship

http://www.w3.org/pub/WWW/PICS/iacwc.htm

[Point] Point

http://www.pointcom.com/

[Psych] Psycoloquy

http://cogsci.soton.ac.uk/~harnad/psyc.html

[QualGuidel] Quality, Guidelines & Standards for Internet Resources

http://coombs.anu.edu.au/SpecialProj/QLTY/QltyHome.html

[Springer] Springer Heidelberg

`http://www.springer.de/`

[Spr_ejourn] Springer Electronic Journals

`http://science.springer.de/e-journals.html`

[SSP_Arch] Full-Text Archives of Scholarly Society Serial Publications

`http://www.lib.uwaterloo.ca/society/full-text_soc.html`

[UCSTRI] Unified Computer Science TR Index

`http://www.cs.indiana.edu:800/cstr/`

[UmichClear] Clearinghouse for Subject-Oriented Internet Resource Guides

`http://www.lib.umich.edu/chhome.html`

[UmichLib] The University of Michigan Library

`http://www.lib.umich.edu/libhome/`

[Umich_rate] Clearinghouse: Information: Ratings System

`http://www.lib.umich.edu/chouse/docs/ratings.html`

[Yahoo] Yahoo!

`http://www.yahoo.com/`

[W3VirtLib] The WWW Virtual Library

`http://www.w3.org/hypertext/DataSources/bySubject/Overview.html`

[W3VL_LOC] The World-Wide Web Virtual Library: Library of Congress Classification

`http://www.w3.org/hypertext/DataSources/bySubject/LibraryOfCongress.html`

[WebCrawl] Webcrawler

`http://www.webcrawler.com/`

Articles

[AAPrep] Scovill, Linda, Librarians and publishers in the scholarly information process: transition in the electronic age: a report from the Joint Working Group on Professional and Scholarly Information in the Electronic Age; under the sponsorship of the Council on Library Resources and the Professional/Scholarly Publishing Division, Association of American Publishers, Inc., 1994

`http://arl.cni.org/clr/Frontmatter.html`

[ACM_Copr] ACM Publications Board, ACM Interim Copyright Policy

`http://www.acm.org/pubs/copyright_policy/`

[ACM_Epub] Denning, Peter J., Rous, Bernard et.al., The ACM Electronic Publishing Plan

`http://www.acm.org/pubs/epub_plan.html`

[Courtois_96.05] Courtois, Martin P., Cool Tools for Web Searching: An Update, Online Vol. 20, No. 3, May/June 1996, pp. 29-36

[Ginsp_96.02] Ginsparg, Paul, Winners and Losers in the Global Research Village, conference paper given at UNESCO HQ, Paris, 21 February 1996

`http://xxx.lanl.gov/blurb/pg96unesco.html`

[Green_94] Green, D.G., A Web of SINs - the nature and organization of Special Interest Networks, 1994

`http://www.csu.edu.au/links/sin/sin.html`

[Hitchc_95.01] Hitchcock, Steve, Carr, Leslie, Hall, Wendy, A survey of STM online journals 1990-95: the calm before the storm

`http://journals.ecs.soton.ac.uk/survey/survey.html`

[Ingold_94.07] Ingoldsby, Tim, AIP's Applied Physics Letters Online: Coming in January (originally in: Computers in Physics, Vol. 8, No. 4, pp 398-401, July/August 1994)

`http://www.aip.org/epub/ingold.html`

[Koch_96.03] Koch, Traugott, Internet search engines, paper given at the first INETBIB-Conference, University Library Dortmund, 11 March 1996

`http://www.ub2.lu.se/tk/demos/D09603-meng.html`

[Odlyzko_95] Odlyzko, Andrew M., Tragic loss or good riddance? The impending demise of traditional scholarly journals

`ftp://netlib.att.com/netlib/att/math/odlyzko/tragic.loss.Z`

[Ok-Donn_95] Ann Okerson, James O'Donnell, Scholarly Journals at the Crossroads: A Subversive Proposal for Electronic Publishing, An Internet Discussion About Scientific and Scholarly Journals and Their Future, 1995

[Roesch_95] Roescheisen, Martin, Mogensen, Christian, Winograd, Terry, Beyond Browsing: Shared Comments, SOAPs, Trails, and On-line Communities, Proceedings of The Third International World-Wide Web Conference, Darmstadt 1995

`http://mjosa.stanford.edu/diglib/pub/reports/brio_www95.html`

[Steinberg_96.05] Steinberg, Steve, Seek and ye shall find (maybe), WIRED May 1995, pp. 108-14, 174-82

[Weibel_94] Weibel, Stuart, Miller, Eric, Godby, Jean, LeVan Ralph, An Architecture for Scholarly Publishing on the World Wide Web,

`http://www.oclc.org/oclc/research/publications/weibel/web_pub_arch/web_pub_arch.html`

[Zorn_96.05] Zorn, Peggy et.al., Advanced Web Searching: Tricks of the Trade, Online Vol. 20, No. 3, May/Juni 1996, pp. 14-28

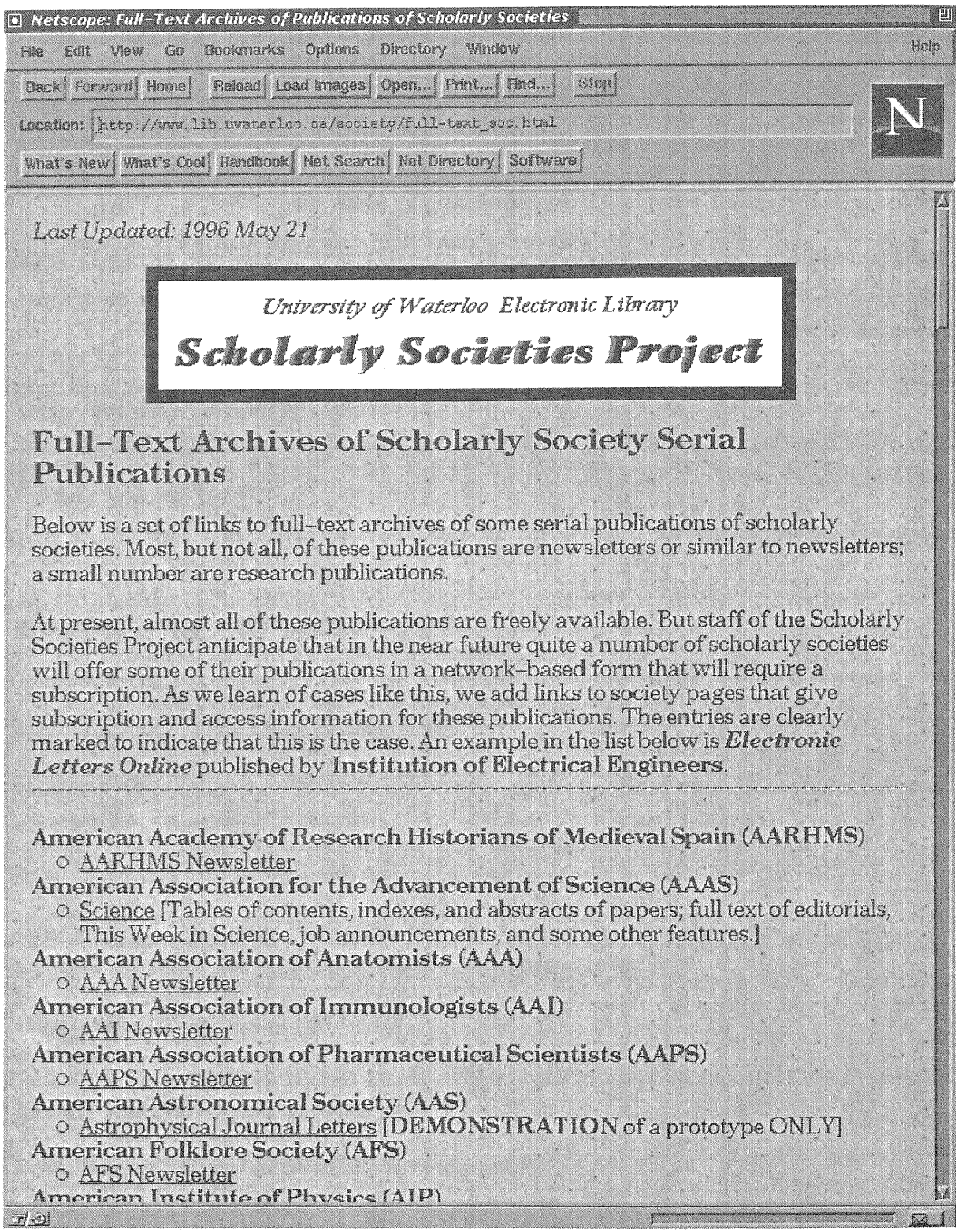

Fig 1 Societies which publish online journals (Univ. of Waterloo)
(図1 電子ジャーナルを発行している学会のリスト (ウォータールー大学))

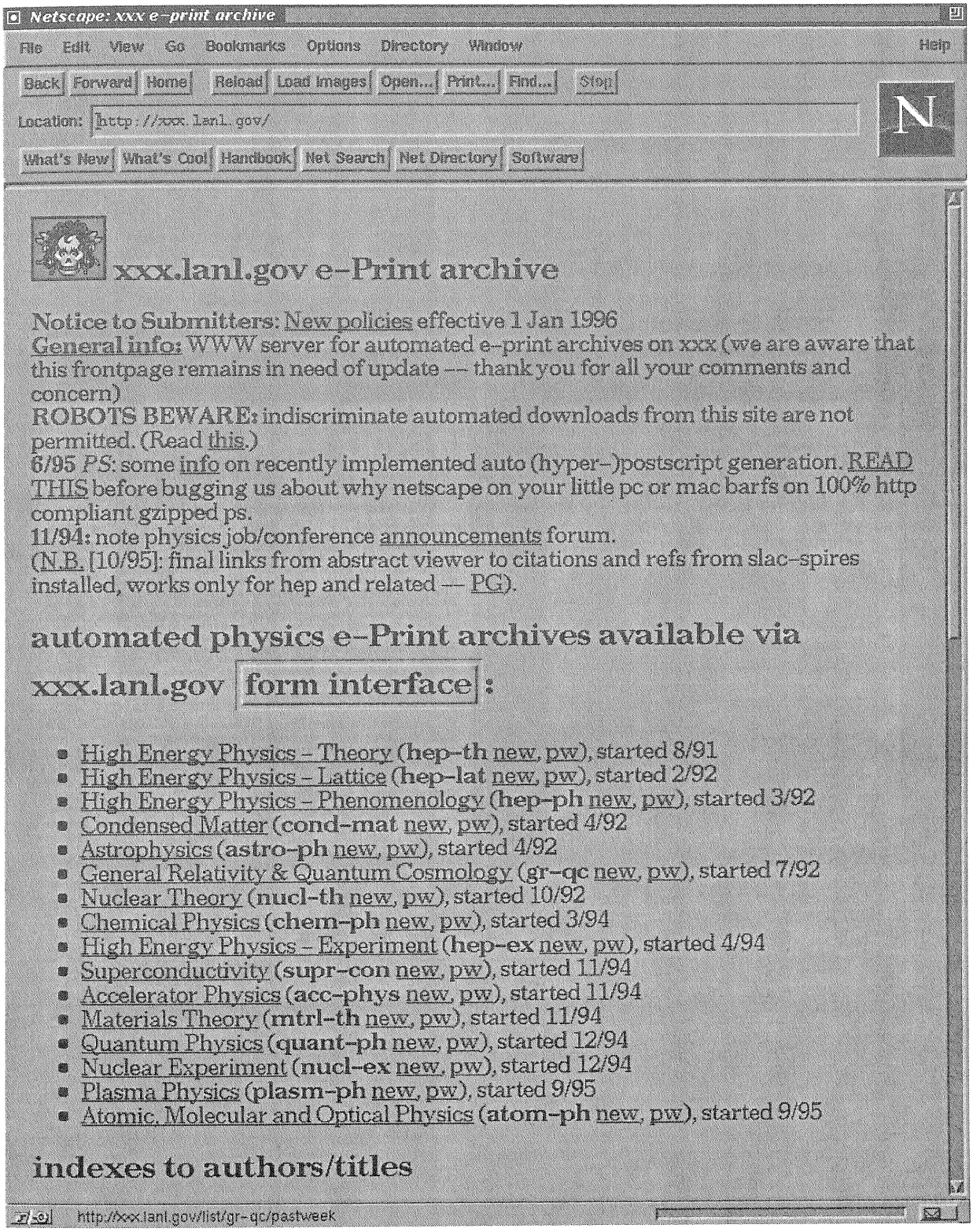

Fig 2 Physics e-Print archive
(図 2 e-Print 物理論文アーカイブ)

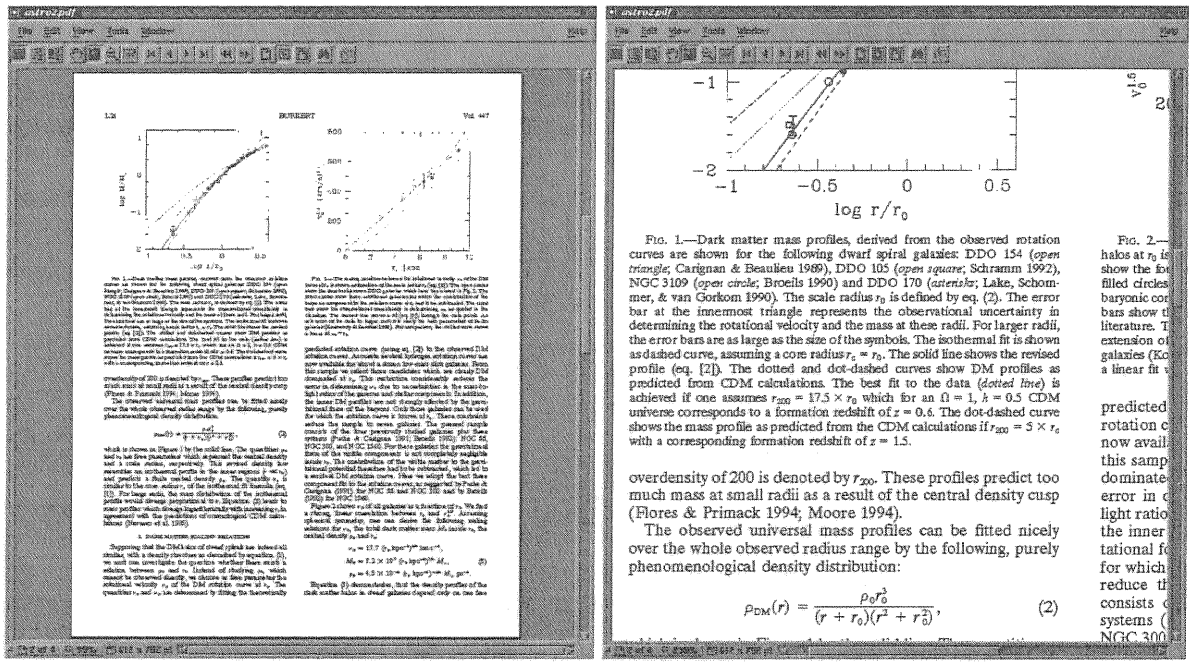

Fig 3 Sample of PDF text image (right: zoomup of partial page)
(図 3 PDF による文書イメージの例 右は一部拡大したもの)

Fig 4 Grouplense — collaborative filtering service
(図 4 Grouplense — 共同作業による情報選択)