materials (or at least the removal of these materials from an electronic information archive) and in other cases change indexing terms in such a way as to distort the functioning of traditional bibliographic access apparatus.

Access to the historical record is not merely an issue of ensuring access to the primary material. In a very real sense, as already discussed, the coverage of the abstracting and indexing services defines the literature of a discipline for many information seekers. But, in almost all cases, abstracting and indexing services began creating computer-processable records in the late 1960s or 1970s. Except for monographs covered in library online catalogs the literature prior to those dates is inaccessible through computer-based retrieval tools, and, for all intents and purposes, might not exist in the mind of many library users. In some fields, particularly the humanities and some social sciences, programs will have to be established to make the remainder of the scholarly record accessible in electronic form and thus place it on an equal basis with the recent publications that are abstracted and indexed by electronic databases. Considerable work is needed in establishing priorities for investment in both the abstracting and indexing of older print material and the conversion of the source material itself to electronic form; these priorities must consider both the mandates of preservation (creating electronic versions of material that is currently deteriorating because it was printed on acid paper, for example) and the programmatic demands of the scholarly communities that will use the materials.

7. The Impact of Micropublishing, Narrowcasting, and Information Feeds

Publishing is becoming increasingly fragmented under the dual pressures of advertisers and readers. For advertisers, a given publication venue becomes increasingly attractive to the extent that it can offer the advertiser a very specifically focused, homogeneous readership-families in the San Francisco bay area that make over $50,000 per year, that are employed by service industries, that collect stamps, and that are shopping for their first house. From the perspective of the reader, overwhelmed by the ever increasing flood of published information, publications that are highly specific to the subscribers interests are of much greater value than more general periodicals. Improvements in the technologies of composition and publication have facilitated this fragmentation to attract readers and advertisers. It is quite common today to see mass market periodicals published in a large number of regional and industry-specific editions; newspapers now offer a wide range of regional editions. Indeed, today's regional editions are composed out of common article databases, but the interest templates are established by region, industry or other criteria rather than individually. The inexorable march of technology seems to point towards ever greater specialization to the audience, ultimately all the way down to the individual reader obtaining a custom product; this trend is manifested in developments that range from the experiments conducted at the MIT Media Lab in the composition of "personal newspapers" based on filters to newsfeeds [Brand, 1987]. Apple's experimental prototype Rosebud system[44] which allows users to define and operate software agents under the metaphor of "reporters" that cover specific information sources [Kahle, Morris, Goldman, Erickson, &

[44] The information access model pioneered by the Rosebud project has recently resurfaced in the commercial AppleSearch product.

Curran, 1992b] or cable television narrowcasts (for example, the slotting of five minutes of local news into a CNN broadcast every hour). Other variations on this model are already coming into large scale use within corporations; for example, in financial and securities markets that are highly sensitive to news, real time news feeds are being licensed for distribution over corporate local area networks, and these news feeds are filtered to allow near-instantaneous delivery of relevant stories to the appropriate individuals very rapidly [Belkin & Croft, 1992; Goldberg, Nichols, Oki, & Terry, 1992; Marshak, 1990].

These trends create enormous problems for libraries, which will only become worse as more information is cast into electronic forms suitable for personalized retrieval. In the case of print, it is becoming very difficult to track what has appeared in the published literature, and who might have been aware of its appearance. Different libraries hold different editions of newspapers and magazines. Abstracting and indexing services do not typically cover all of the various regional editions of a publication; rather, they select one (sometimes without much consideration, and without being very clear about which edition they have selected) for indexing. As we move beyond print editions to the databases from which the specialized print editions (and future personalized extractions) are generated, the situation changes again. It now becomes possible for a library to obtain or purchase access to what is in essence the composite intellectual contents of all of the various editions, but the researcher coming to these databases years later may well loose any understanding of how items from these databases were selected (or, sometimes more to the point, not selected) and presented to any given audience or what the impact might have been on readers. The selection of available materials into a given edition represent a value judgment about the importance of specific subjects in time and space, and the record of this judgment is of critical importance to researchers. The key questions, as we attempt to mine these databases for research, will be what material a given reader of a given print edition learned from that print edition, and what material would likely have been selected from a given database at a given time by someone filtering the database with a given interest profile. To make matters worse, the filtering typically occurs (at least today) close to the end user rather than the publisher, and there is a great diversity of filtering tools ranging from rather simplistic keyword matching all the way through sophisticated research tools that perform semantic and linguistic analysis to identify material that may be relevant.

There is also a serious problem with the accuracy of citations while we remain in a transitional stage between print and electronic databases. n the fully electronic world, one might simply reference article ID X (added on such and such a date) in the Wall *Street Journal* article database. But today, this would most likely be referenced by its title, and the date and issue number of the issue that the article appeared in, without providing the critical bit of additional information—was it the East Coast, West Coast, European, or Far Eastern edition?

The recombinant nature of materials that are maintained in electronic databases but presented to the reading public through specialized printed vehicles is not limited to newspapers and magazines. McGraw-Hill has embarked on a program called PRIMUS in which they supply databases of articles that can be combined into course readers, with the inclusion of optional commentary or additional material by the compiler of the

course reader. Here again it is becoming increasingly difficult to determine the provenance of material, its timeliness, or the conclusions that the reader might have drawn from it based upon context.

The ready availability of desktop publishing (now being combined with transmission via fax or electronic mail across telecommunications facilities) has led to an incredible proliferation of very narrow audience newsletters; many of these have very high time value, very high subscription costs, and sometimes rather high impact on the communities they serve. Yet to those outside of these select subscriber communities, these publications are almost invisible. They are not available in libraries, except occasionally for specialized corporate libraries (which often will not circulate the material through interlibrary loan), and material in them is not indexed in commercial abstracting and indexing services that researchers would use to try to obtain information about a subject, or to reconstruct events after the fact.

A surprising parallel can be drawn to the infamous "gray literature" of technical reports in some areas of science and engineering; these are poorly indexed, difficult to obtain documents that play a key role in circulating information within an insider community, but are not readily accessible either through libraries or through the bibliographic control and access apparatus of abstracting and indexing databases. Yet, for an active researcher, they are invaluable, and the information propagated through such technical reports can have an extraordinary impact on a scholarly community. The contents of the technical reports will appear in the traditional, "archival" published literature only years later in many cases, and often in an edited (reduced) form due to page limitations in the archival journals. In some disciplines, important work that is described in technical reports does not always make its way to the mainstream literature, with the authors simply viewing it as too much trouble to manage a manuscript through formal submission, peer review and publication when the material is already somewhat stale and dated to the authors. In some academic disciplines, it is probably not too strong a statement to argue that were it not for the need to publish in traditional archival print venues for tenure and promotion purposes, most of the effort would go into producing material that would be part of the gray literature. Increasingly, in part as a recognition of the importance of such material in scholarly communication and the diffusion of research, such reports are being made accessible through the Internet from file transfer (FTP) or GOPHER servers, most often through initiatives at the departmental level. Of course, this leads to an essential literature that is very hard to obtain access to, particularly for those who are not comfortable with the relevant technologies, and the long term accessibility of this literature is questionable, since there is today little institutional commitment to ensure continued availability of it; partially in a response to these trends ARPA is funding a major project to improve access to technical reports in computer science through the network. 45

[45] If anything, the scope of the gray literature is becoming larger and more confusing as we move into an electronic environment when anyone can make material directly available for anonymous FTP from a personal workstation or departmental storage server. Not only **are** members of **the academic community mounting technical reports but in some cases also the text of material that they have published,** often in violation of the transfer of copyright agreements that they have executed with the print publishers of the work as a condition of publication. Thus far, this has occurred only at the level of individual authors and not on an institutional basis and to the best of my knowledge the print publishers have not attempted to enforce their control of the rights with these individuals, perhaps feeling that it is not worth the expense or fearing that such an action would mobilize the community of academic authors to pay more serious

Over the past few years, the research library community has begun to devote resources to improving access to the technical report literature as part of a recognition of its importance. Several major research libraries have developed extensive technical report collections, and Stanford University has developed an extensive bibliographic database of technical reports which is available on the Internet. Yet the newsletter literature remains almost inaccessible to the academic community and the general public.

While somewhat outside of the primary focus of this paper, mention should also be made of real-time information feeds that go beyond human-generated intellectual property such as newswires. Other types of real time information feeds are beginning to appear on the network; these information feeds, just like newswires, can be analyzed by personal workstations to sift out events of interest. If ever knowledge was power, these applications illustrate the axiom; in some cases, the advantage of timely knowledge goes beyond financial or professional gain to matters that are literally life and death. The ownership, access to, and integrity of these resources is of potentially critical importance, and may in future become a central public policy issue. Consider the following examples of information feeds that might pass over the network:

- Newswires. These are clearly intellectual property (representing reporting of news events) yet timely access to news may permit an individual to obtain substantial financial or professional gain. Related to newswire feeds will be a wide range of real-time audio and video telemetry of current events of interest, similar to what is sometimes found on the C-SPAN cable television network or carried on current experimental services such as Internet Talk Radio. The intellectual property rights involved in these audio and video feeds is far less clear; in some cases the events are public (for example, Congressional hearings) and in many cases they are simply captured through fixed video cameras and microphones, thus involving little creative work.

attention to copyright transfer agreement terms. Often, it is extremely hard to tell precisely how the electronic version available from the author is related to the printed publication; in some cases, the electronic version does not even provide a citation to where the work appeared in print. And, of course, the electronic version is often available months or even years prior to the availability of the published printed paper. Another interesting area of expansion of the gray literature is doctoral dissertations. Most universities use University Microfilms Inc. (UMI) as a means of ensuring general availability of theses, and request (sometimes essentially require) doctoral candidates to file their dissertations with UMI as part of the process of filing the dissertation. The UMI agreement gives the author non-exclusive print rights, including the right to publish articles or books based on the thesis, but reserves *exclusive* electronic distribution rights to UMI. Yet theses, or versions of theses that are issued as technical reports by academic departments are starting to appear for public access as well. Again, it is difficult to tell if these are precisely the same as the formal filed theses; the print versions that one obtains from UMI normally include the signed cover sheet and are identical to the versions filed with the university. H should also be noted that while the incidents discussed here of making dissertations (or variants of dissertations) available electronically are isolated individual cases, there are also institutional level discussions going on between a number of universities and UMI under the auspices of the Coalition for Networked Information about how to make dissertations available electronically through the network and what (if any) UMI'S role should be in such an enterprise. These discussions might well result in a situation where UMI serves as an electronic publisher for dissertations on behalf of the university community (in print, UMI is a very cost effective means of providing access to dissertations, with a typical dissertation only costing about $35); alternatively, it might result in universities negotiating for changes to the UMI copyright transfer agreement and mounting dissertations themselves.

- Stock or commodities prices, or other financial data (for example, purchases of rare coins). It is much less clear that this is intellectual property rather than brute facts, but again timely access to and analysis of such information can offer an individual substantial advantage. Currently, one can obtain subscriptions to the stock market trading “ticker” (at considerable expense).

- Time synchronization information. The US Government, as a public service, offers very precise time information suitable for setting clocks in a distributed computing environment so that multiple computers can operate on a common time scale. (These efforts are supported by NIST, based on programs that date back to when NIST was the National Bureau of Standards). A related service is the Global Positioning System (GPS) satellite network, which allows a user to locate the position of a receiver with a tremendous degree of accuracy. GPS was deployed to support military applications, but has extensive civilian uses that include search and rescue, navigation (for planes, boats, and even potentially automobiles or individuals on foot). Currently, such information is publicly available, and is of substantial value; will private commercials services have a role here in future?

- Weather information. Much of this is collected from sensors (obseving stations, satellites, etc.) financed through public funds. Yet subscriptions to weather information are in many cases through private information services (though the University of Michigan, for example, makes information derived from such private services available to the general public across the Internet through the Weather Underground service). In some cases this information is merely interesting (e.g. whether it will rain today); in other cases it is of financial value (implications for commodities prices). But in a few cases—tornado watches, flood warnings, and the like—it is a matter of great importance to recipients, if they can obtain the information in a timely fashion and act upon it.

- Surveillance data. We are awash in digital imagery ranging from data generated from satellites (either government or commercial, such as the French SPOT system) through output of surveillance cameras in our workplaces and homes. Again, awareness of such information can have value ranging from insight into commodities or securities investment through convenience-traffic congestion information, for example—to personal safety.

- Earthquake warnings. Earthquake shock waves propagate from an epicenter rather slowly compared to the near speed-of-light propagation of information across copper or fiber optic trunks. Some states, such as California, have sensor networks deployed which could, at least in theory, propagate information about the occurrence and epicenter location of earthquakes across the Internet in such a way that locations that will be hit by a major earthquake might obtain as much as 60 to 90 seconds advance notice of the event. This is information which, if identified and acted upon in a timely fashion, could save not only a great deal of money (for example, by “safeing” processes ranging from parking heads on disk drives or stopping elevators in a graceful fashion at floors through shutting down industrial operations such as chemical refineries) but could also save lives. It is unclear who would own such information, or could provide it to the network.