ELECTRONIC PUBLISHING, VOL . 6(2), 65–66 (JUNE 1993)

Editorial

The observer of the research scene will have noticed increasing discussion of the concept of a 'Digital Library' in recent months. In the United States, a funding effort led by the National Science Foundation holds the promise of supporting several large-sized demonstration projects and consequently has attracted the attention of academics. Within the last few weeks, in the UK, the Follett Report has also recommended that substantial research funding be made available in this area. With this kind of interest in 'the library of the future', on both sides of the Atlantic, it seems like a good time to reflect on what needs to be done.

The first requirement will be for flexible, high-bandwidth communications. Already we see increasing access on a global scale to the Internet, and its consequent popularization in the general press, coupled with the development of enabling applications such as the United States' National Center for Supercomputing Applications' Mosaic software. This has given many people their first introduction to browsing through the already-existing, but clearly primitive, large-scale distributed digital library that has arisen in an *ad hoc* fashion from the many individually-developed components found throughout the Internet.

There are some who will claim to have invented the idea of a narrowly-defined Digital Library and there are others who will trace the history back at least as far as Vannevar Bush's Memex, which he described in 1945 in an article in *The Atlantic Monthly*. In either case it seems relevant to retain, or at least consider, the connection with the traditional library. A library is a repository, and consequently implementation of a digital library revolves around choosing the kinds of objects that will be stored in the repository, designing the mechanisms for presenting and manipulating those objects, defining the structure of the repository itself, providing the means to permit access to objects, and creating the techniques that enable location and selection of particular objects from within the collection. However, focusing entirely on the collection of a traditional library — categorizing the kinds of media stored there, studying the construction of shelves to hold the collection, and designing a card catalog (perhaps carefully specifying the kind of wood used to build the card cases) — fails to capture much of the value of the library to its patrons.

An important role of a library in a research community is to support collaboration, both across time between a scholar and his predecessors, through the intermediary of the knowledge encapsulated in the collection, but also collaboration at the same time between contemporary scholars, with the library's artifacts serving as a catalyst for discussion. The Digital Library presents opportunities to augment the support for collaboration by adapting and adopting the methods studied by workers in the field of Computer Supported Cooperative Work.

Also of importance in the smooth running of a traditional library are the techniques and standards that have been developed to define and maintain the policies that establish the characteristics of the information space. Such policies include questions of what to obtain for a collection, what to retain in it, and how to categorize and catalog a

CCC 0894–3982/93/020065–02 *Received 12 December 1993*

collection's objects. A characteristic of current Digital Library initiatives is that they represent collections of significant size. Handling the issues of management of large scale collections of information plays a central role in the success of any library, be it digital or traditional.

It is perhaps this characteristic of scale that helps to distinguish work appropriate for application in a Digital Library. The argument is often made that certain techniques or implementation will fail to scale up, either for performance or for conceptual reasons. Such techniques, while appropriate for application to small-sized problems, will not be adequate for handling massive collections of data. Conversely it seems reasonable to hypothesize that techniques developed for application in a Digital Library may fail to *scale down* — metaphors appropriate for managing scholarship in the large sense may not be well-tuned for supporting individual, independent, introspective scholarship.

The long-term reader of EP-odd will note the strong overlap between what has been discussed in the journal and the techniques required to implement a Digital Library. Indeed, the reader may note that the question of appropriateness of scale is one that pervades Electronic Publishing as well as Digital Libraries. Our technical reviewers, when evaluating reports of such investigations, generally require specific illustration that the described techniques work in practice — evidence that the techniques have been implemented and have been given a try with their intended user population. The reason for this is that most reviewers have, in some distant corner of their past, encountered at least one idea of theoretical purity and beauty that failed miserably when applied. A requirement that the acceptance of a technique be subject to demonstration of practical experience provides a sensible checkpoint for protecting the journal's integrity.

Investigations involving the management of massive amounts of data require the investment of a significant amount of resources to properly validate the results of the research. Consequently it is encouraging to note the initiatives that are under way to provide the funding needed for the implementation and deployment of realistic Digital Library prototypes. The amount is certainly not insignificant; for example, the United States' National Science Foundation expects to fund a number of projects, each for about $1.2 million for each of four years. The Follett recommendations call for £2 million to be spent over three years for the development of electronic journals, £1 million for the promotion of digitized texts and £1 million for electronic document delivery projects. Clearly, initiatives at this level will favor a conservative management plan and consequently will require a noticable amount of management infrastructure. It is interesting to speculate whether the research program will also take a conservative route, thereby decreasing the level of innovation.

Taken as a whole, the Electronic Publishing community benefits from a mixture of large and small research projects as well as a mixture of long-lived and short-lived projects. To date it has been difficult to find many projects based in Universities that address the issues of size, so the effect of larger funding intiatives will be monitored with interest. However in support of innovation, it is critical to continue to support smaller, more flexible projects, while keeping carefully in mind the limitations of such efforts.