Introduction

THOMAS A. PETERS

The idea of a theme issue on the topic of assessing digital library (DL) services immediately raises at least two red flags. First, have DL services evolved sufficiently to be amenable to a sustained organized assessment effort? In the last several years, online library services have grown by leaps and bounds, but we were starting almost from scratch. Are baseline data and best practices ready to be harvested? As a profession, are we ready to gather and gain sustenance from the harvest? The answers to this first set of questions depend in part on one's philosophy of assessment. Some may argue that planning for assessment should be one of the activities of the pre-planning stages of any library service project or program, while others may argue that a sound assessment plan can emerge only after the project or program to be assessed has been in existence for some time.

Second, what types of assessment models and methods need to be adapted, adopted, or created for this field of assessment? At first glance, it appears that it would be logical to rely on the mature field of assessing public services in physical libraries as the template for assessing DL services. On closer examination, however, we notice some pronounced differences between the two service environments. Online information environments can be structured differently than physical information environments, and online information environments can contain multiple structures simultaneously, unlike physical information environments that are informed by pre-use structural and organizational decisions made by architects and librarians alike. Also, in general, humans are becoming accustomed to pursuing their information needs in online environments.
It is doubtful that reliable patterns of user needs, wants, expectations, and behaviors have become settled, particularly when we consider online services.

The articles in this issue attempt to lay a foundation for the potentially large diverse field of assessing digital library services. Borgman, Gilliland-Swetland, Leazer, Mayer, Gwynn, Gazan, and Mautone provide an overview of the evaluation project related to the development and deployment of the Alexandria Digital Earth ProtoType (ADEPT), a digital library of geographical information designed to be used in conjunction with undergraduate education. One goal of ADEPT is to expand on the testbed architecture developed for the Alexandria Digital Library (ADL) in a way that focuses on users, use, and outcomes. The authors concentrate on how DLs can facilitate the integration of information technologies into campus-based instruction. They note that, if DLs are to be assessed in terms of learning outcomes, many fundamental research design questions need to be addressed in rapidly changing information and instructional environments. Their thesis is that DL services "will contribute positively to undergraduate instruction and to student learning of scientific processes." Their chief interest is to understand how the use of digital libraries can promote thought processes associated with problem domains. They want to assess the "cognitive consequences" of participating in an ADEPT environment. Thus the goals of their assessment program go far beyond the concept of a digital library as a self-contained online system of information and information services. The type of assessment examined in this article concerns learning outcomes and, more fundamentally, thought processes. They note that the browsing capabilities of DLs can aid students' question-asking, one of the five skill sets needed to engage in scientific thinking in geography. The authors assert that "digital libraries are more than storehouses of information; they should be aids to the question-asking, information-gathering, information-organizing, information-analyzing, and question-answering processes of users." They suggest that formative evaluation is a useful approach in most current DL situations. The design team and the evaluation team have undertaken an iterative and collaborative approach to their tasks. Three types of evaluative studies (employing both qualitative and quantitative methods) are being conducted to assess ADEPT: classroom-based studies, laboratory studies, and system use studies. Based on their initial classroom-based studies, the authors realized that ADEPT modules need to be flexible, adaptable, and relatively small in scope.

Carter and Janes present the results of an exploratory study that attempts to establish a methodology for the unobtrusive automatic analysis of an online reference service. Since the Internet Public Library's (IPL) digital reference service began in March 1995, over 40,000 questions have been handled. IPL's reference service is an educational enterprise, and
many of the reference service providers are graduate students still in the process of learning reference techniques and service styles. The authors note that, because reference services delivered over the Internet currently are mediated primarily in a textual way, we have new ways of examining online reference activities. For the study reported here, over 3,000 reference interactions occurring during the first quarter of 1999 were analyzed. The IPL currently offers two forms for asking questions—a general form and one designed specifically for youths. Only 4 percent of the questions were submitted on the youth form, and 26 percent of all questions arrived as free-form e-mail questions. Methodologically, the authors are interested primarily in what sorts of insights can be gleaned from the cluster of interactions using automatic means. Consequently, no content analyses were undertaken nor were patrons queried directly regarding their impressions of the online reference service. They found that the median turnaround time to answer questions was just over two days, and approximately 20 percent of the answers caused users to send thank-you messages. People who submitted questions related to library science and music exhibited a greater tendency to submit “thank you” notes. There was frequent disagreement between the subject areas of questions as identified by the asker and as determined by the reference service provider. The authors note that, because many users of online reference services have difficulties articulating what their question is about, automated disintermediated reference services face a substantial challenge. This is a fascinating study in a nascent field. It raises many interesting questions about how to interpret, categorize, assess, and improve online reference services. The authors conclude that “when designing a reference question intake form, librarians should consider not only what they will need to answer the question, but also what sort of automatic data analysis they may wish to do in the future.” This article contributes significantly to our understanding of the possibilities for unobtrusive analysis and assessment of online reference interactions.

Gorman, Ash, Lavelle, Lyman, Delcambre, Maier, Weaver, and Bowers implicitly suggest that any attempt to create and evaluate sustainable, useful DL public services must be premised on a deep understanding of how real people look for, make sense of, and manage information in real-life situations. For example, they explore how experts often create and use “bundles” of information (i.e., organized collections of highly selected information) to solve problems and maintain current awareness of situations. Bundles of information are created to help perform specific tasks. Digital libraries should incorporate computer-based tools for creating and managing bundles. This article describes aspects of a larger DLJ-2 (Digital Libraries Initiative, phase 2) funded project, “Tracking Footprints in an Information Space: Leveraging the Document Selections of Expert Problem Solvers.” When health care professionals (e.g., critical care nurses,
Resident physicians, attending physicians, pharmacists, and medical ethicists are focused on gathering information that may help to solve a specific medical condition, they make explicit choices about which items to ignore and which items to examine more carefully. In particular, messy bundles are valued not only for their convenience, but also for their immediacy, portability, disposability, and flexibility. The authors note that, for many information seekers, finding a satisfactory solution that suffices is preferable to devoting substantially more time, attention, and effort to finding the optimal solution. They also question whether these bundles can be reused by the same person/team or others. The authors also explored how experts combine information into high level scripts that trigger retrieval of additional details from memory.

If a satisfactory set of DL services were being designed for this population, what would these services have to do? How could they improve on the paper-based versions currently being used? The authors also indicate an intriguing behavior that could be called ignorance, defined as the assertive decision and action to ignore certain information objects.

For Greenstein, a digital library "mediates between diverse and distributed information sources on the one hand and a changing range of user communities on the other." He makes a strong case for services as the distinguishing characteristic of a digital library. Because, in a digital environment, a library assumes responsibility for configuring access to a world of information, a digital library is known less for the collections it owns than for the networked information space it defines through its online services. Because the digital library is "evolving as the library's defining function and as such is developed with a view to its financial and organizational sustainability," assessing a digital library is a high stakes endeavor. Greenstein suggests that the emerging business-to-business economy for networked environments could be mimicked to supply a class of infrastructural DL services that are more effectively mounted on an institutional or even cross-institutional level. Lest we fall into the belief that assessing DL services is solely a professional prerogative and activity, Greenstein reminds us that there is a distinctive need for benchmarks that help users evaluate DL collections and services. Greenstein's article also reminds us of the symbiotic relationship between digital collections and digital library services.

Marchionini describes the multifaceted and longitudinal (beginning in 1987) evaluation efforts related to the Perseus Digital Library (PDL) (http://www.perseus.tufts.edu), an evolving digital library for the study of aspects of the ancient and modern worlds. He notes that a DL is a marriage between the cultures of physical libraries and the sometimes disparate cultures of computing and telecommunications. Marchionini notes that the problem of assessing and evaluating digital libraries is one of assessing complex adaptive systems. The ultimate goal of evaluation is
to assess the impacts of libraries on patron's lives and the larger social milieu. Marchionini cautions that it is important to assess the impact of a DL over a long period of time. He notes: "Perhaps the most important long-term developments are changes at organizational levels such as departments and schools and the emergence of a community of practice that leverages and advances the PDL." He reminds us of the important distinction between evaluation as a research process and evaluation in the product testing and system efficiency sense. Marchionini defines evaluation as "a research process that aims to understand the meaning of some phenomenon situated in a context and the changes that take place as the phenomenon and the context interact." The ongoing PDL evaluation program depends primarily on educational evaluation. Four types of data collection methods are used: observation, interviews, document analysis, and learning analysis. He notes that assessing interactivity remains a significant challenge as we try to understand browsing and other interactive behaviors in online environments. Marchionini cautions that the motivations and styles that students bring to learning tasks affect the effectiveness of a digital library. Although the Perseus Digital Library focuses on content and emphasizes self-directed learning, the implications of this excellent longitudinal evaluation program on the assessment of DL public services in general are obvious.

Peters suggests that, at the present historical stage of the development of DL services, particularly online reference services (ORS), it may be more fruitful to concentrate on "meta-assessment" activities than on assessment activities proper. Meta-assessment is defined as the deliberate examination of the elements, basic conditions, and needs of a service program that transcend particular instantiations of a particular thing. Meta-assessment does not entail assessing particular programs but rather the conditions (if any) under which all online reference services must exist. It occupies the conceptual space between the philosophy of reference service (i.e., the examination and articulation of first principles) and the assessment of a particular reference service program. Peters also raises questions concerning the impact of "rogue" reference services (i.e., ORS that are not affiliated with any particular digital library) on the process of meta-assessment. Peters concludes that, although the widespread recurring assessment of specific ORS may be a few months off, the window of opportunity for an optimally effective meta-assessment of ORS in general may be closing. Now is the best time to engage in meta-assessment activities because expected patterns and modes of online reference service have not yet been established, and because the distance between theory and practice is at its perigee.

Noting that evaluation has not been a conspicuous activity in many DL projects and programs, Saracevic concentrates on clarifying the definitions and possible taxonomies related to DL evaluation programs.
Saracevic warns that evaluation at these early stages of DL development and evolution could have dangerous stifling effects. Among other things, Saracevic suggests that the conceptual state of the art of DL evaluation may be insufficiently developed at this point in time to be widely practiced. He advocates a systems approach to evaluation, defining evaluation as an appraisal of the performance or functioning of all or part of a system in relation to some articulated objectives. Clear specifications regarding construct, context, criteria, measures, and methodology are required for any evaluation of digital libraries. Any DL evaluation project or program must specify clearly what elements (e.g., collections, access, services, costs, etc.) are being evaluated. The levels of a DL evaluation can range from a broad social level through the interface to the engineering, processing, and content supporting the system. Saracevic notes that, to date, evaluations of digital libraries have not been conducted on more than one level. Criteria from the evaluation of traditional libraries, traditional information retrieval systems, and traditional HCI studies can and must be adapted for DL evaluation projects. Uniformity across DL systems and persistence over time may serve as additional evaluation criteria. Saracevic offers multiple perspectives for conceptualizing the structure and purpose of DL service assessment projects and programs.

Seadle provides an anthropological perspective on the National Gallery of the SpokenWord, an NSF-funded DLI-2 project centered at Michigan State University. He argues that any attempt to assess the worth and outcomes of a DL program must be preceded by an attempt to understand the people (including their “micro-cultures”) involved, ranging from the co-principal investigators of the grant-funded project to the real and anticipated end-users. A micro-culture denotes units of shared meaning as small as academic professions, university departments, and interest groups. Most DL users have links to, and are molded and influenced by, various micro-cultures. Seadle notes that the meaning and goals of a DL often are elusive. The texts emanating from a DL project, such as the original grant proposal, often do not fully reveal the goals and meaning. Seadle suggests that the methods and mode of cultural anthropology can provide the intellectual foundation upon which informed choices can be made concerning the sample population, survey design, and focus group selection for the assessment component of the project. Seadle concludes, among other things, that the imprecisions of meaning lie at the core of the evaluation issue regarding DL services. The different micro-cultures involved in this DL initiative make different assumptions about the composition, needs, and desires of the end-user population. The introduction of cultural anthropology to the DL assessment process is a beneficial message for librarians.

The question concerning the design and deployment of projects and programs to assess digital library services contains many discussion and
decision points, ranging from the relentlessly practical to the purely theoretical. The emergence and maturation of DL services over the past few years has been impressive. We need not be ashamed that some of this development has been fueled by adopting software originally designed for other purposes, such as Web-based call center services. In the emerging field of online reference services, the adaptation of CRM (customer relationship management) software to the online reference encounter has been a tremendous liberating breath of fresh air.

As one would expect, this group of articles raises more questions than it answers. Online environments in general, and DL environments in particular, continue to evolve at a rapid pace. Imagine trying to provide reference desk service in a physical library while the architect changes her or his mind about the basic disposition of the edifice. Imagine trying to serve a populace that is still formulating its basic expectations about the environment in which the service is provided. These are the challenges faced by the profession as we develop assessment programs for digital library public services.