As a first step towards the design of the testbed for the Stanford Integrated Digital Library Project, we are working to create an explicit shared understanding of our key assumptions. These assumptions lay a foundation for choosing our system's framework (the architecture, platforms and tools); for the detailed design work needed to build interfaces, protocols and program specifications; and finally for the implementation of code. They are not decisions about specific design choices, but statements of criteria and concerns that will shape the decisions we make in defining and designing the project.
1. NATURE OF THE PROJECT

1.1. We are designing and placing into operation for evaluation an experimental research vehicle to demonstrate interfaces and mechanisms for future digital libraries.

Although our overall vision is the integration of a wide range of materials in a system that can be used by many people, we are not planning to produce a product for general use in the initial four years. The most effective use of university research resources is to develop advanced prototypes that explore concepts and mechanisms that can later be incorporated into production-level software for widespread distribution and use. We will build a working system -- our research strategy requires the development of a working testbed, in order to really test out the ideas. This testbed may be a "throwaway" in which design tradeoffs have been made to maximize the potential for exploration, rather than to ensure robustness and coverage beyond what is needed for the testing.

1.2. Our work will predominantly have a technology focus.

In exploring possibilities for the digital library of the future, we will be looking primarily at the technological mechanisms that underlie new forms of integration. We will work closely with our library and publishing colleagues to understand and seriously account for the situations of use. We hope and expect that our general framework for integrating library contents and services will attract services that explore the broad non-informational characteristics of libraries as well; those processes and interactions that are more community and social. But we will not do extensive research on the social questions of the nature of libraries and their roles in communities.

1.3 We will need to set specific measurable goals and milestones for what we achieve.

In order to evaluate the work, there need to be measures of success that go beyond (although not to the exclusion of) the traditional academic measures of peer recognition of the value of the ideas (reflected in publications, conferences, etc.). We are beginning to shape these goals, and some initial possibilities for what we might have done in four years are:

a. We will have developed protocols and standards that are usable and proved in reasonable scale settings.

b. We will have developed a suite of interfaces and services that can be the "method of choice" for at least some information users in our test collection. That is, they would find our testbed the most valuable tool they have for some significant part of their information using activities.

c. We will have developed a set of "key scenarios" of how people use a digital library, and will have gathered empirical data on how well we can support those scenarios.
These are not an exclusive list, but the starting point for further thinking about how to set our goals.

2. WHAT IS "THE DIGITAL LIBRARY"?

The term "digital library" evokes many different responses. "The library" can be generic (like "the computer") or specific (like "The Library of Congress"). A library has many different aspects -- the physical facilities, the set of "holdings", the collection of services it provides, and the community of users of different kinds. We will develop our project in accord with the following views:

2.1. The "digital library" is more than a collection of digital information items: It is a coordinated collection of services, which are based on collections of materials, some of which may not be directly under the control of the organization providing a service in which they play a role.

Increasingly, as materials move on-line, there will be potential access mechanisms for anyone. The distinction of a "library" will be based on the structure of access to its services rather than on the private possession of specific collections of information. The digital library services include the facilitation of publishing, organizing, and accessing media of all kinds, including print, analog videotapes, photographs, and the like.

2.2. There will continue to be many libraries, not a "unified universal digital library."

These multiple digital libraries will be constituted as separate organizational and financial entities, which in some cases will be based on separate mechanisms and/or separate resources. One traditional library function is to redistribute the costs of services. Public libraries, university libraries, and (for many of their functions) corporate libraries do not charge fees on a per-use basis. There are often special fees for "special" services, but there is a baseline level of service with costs distributed along organizational lines. The sponsoring organization does this for many reasons. Imagine the educational consequences of a university library where students pay by the number of items they check out to do their work. The number of distinct libraries and the policies they provide for access will be determined by social, economic, and organizational boundaries, not informational ones. For this reason, the digital library world of the future will consist of many libraries, not one worldwide unified library.

2.3. The digital library will integrate the on-line information world and the physical world of information objects.

For the foreseeable future a large percentage of important materials will be in non-digital form. The digital library needs to integrate meta-information about these as a major aspect of its services, rather than concentrating only on fully on-line materials. This will be less true for our particular testbed collection, since material about computing tends to be relatively new (created after computers existed) and done by people who are predisposed to use computer-based media. But we don't want to design a library that won't work well for other disciplines and for public non-academic use. What is "digital" is the
3. USERS

3.1. The Stanford DL testbed will support users who are early adapters of technology, who are computer literate and who have an interest in the computing literature.

We intend to build general mechanisms for a wide variety of library uses and users, but our testbed (including the materials and any interfaces we develop) will be designed for users and needs of the kind typically served in a university or industrial research library. These users will use INFObus services to support traditional activities associated with information production, acquisition, organization, storage, retrieval and manipulation.

3.2. Our users span the publishing/end-use spectrum, from authors and publishers to traditional library functions to end-seekers of information.

Much of the on-line information systems discussion in the recent past has focused on the end user's perspective. Successful library tools will put equal emphasis on other parts of the information chain, providing tools that enable people to produce, organize, restructure and disseminate information as well.

4. SERVICES

4.1. The INFObus should be designed to support a wide variety of services, and it will be tested by our building selected in-depth examples based on opportunities for leverage.

It is impossible to predict what kinds of services are needed and what will characterize the successful service. Rather than dictate or guess what will be required, we will provide an environment that allows experimentation and the incorporation of services developed by others, both in the digital library project and in the commercial market. We have identified a core of service types that will be supported (payment, security, retrieval, archiving, etc.) Within these general categories, we will support unlimited numbers and kinds of services. In choosing examples to implement, we will seek to make effective use of our existing experience, collaborations and expertise.

4.2. The key to developing effective library services is to give people quality, not quantity.

The natural forces at play in the networked environment will rapidly make it easier and easier to retrieve all kinds of information from all kinds of sources. The research problem to be solved by our project is one of providing the underlying concepts and tools that make it possible for people to provide mechanisms that add value for users, beyond simply providing access.
4.3. Library services supported by the INFObus should be providable by a relatively large number of people and institutions, with varying degrees of resources and sophistication.

Traditionally there have been a very small number of publishers and librarians relative to the population that uses the information they provide. In the digital environment the boundary between provider and user breaks down, and the mechanisms and tools on the provider side need to be usable by a wide range of people. The INFObus design should facilitate the provision of value-added services by people who do not have large storage resources or highly specialized knowledge or equipment. The level of platform required for authoring, publishing and library services should be within reasonable reach of a large proportion of the user community.

4.4. The INFObus should support a wide range of service models for security, authentication, intellectual property control, billing and payment, etc.

Our collections will include materials that are published, unpublished, protected and in the public domain. At this early stage of digital libraries, many models of control and payment will be built and tested by many different interested parties. Complex legal and social questions are arising about the balance between new technologies and the traditional rights of the research and educational community and of those who produce information. It is not the job of our project to make these decisions, but to provide a service model that is open and can support the management of materials in accord with a wide range of desired policies.

5. CONTENT

5.1. The focus of digital libraries infrastructure should be on facilitating the provision, dissemination, and use of information by institutions and groups.

When considering the overall information context, it is tempting to include all of the different modalities and aspects: desktop video conferencing, real-time text interaction (MUDs, chat, etc.), email, shared screens, etc. Although it is likely that in the long run all of these will be integrated in a unified system, it is well beyond our resources to do anything that would adequately address all the issues. A reasonable focus is to concentrate on asynchronous information (it is provided by some person or group at some time for later use by others) and on information flow which is mediated by information structures (as opposed to direct person-to-person modalities such as personal fax or email).

5.2. We will focus on materials in English.

Although the long-term goal of a digital library needs to be international, we can better focus on the research questions we have chosen without getting entangled in the problems that many other people are already pursuing about multilingual representations and interface support.
5.3. Digital library mechanisms will be used for a wide variety of materials.

The mechanisms should not require that materials be of a certain form, or have a certain degree of "certification" to be included in a digital library. We should provide means for people to declare that certain materials are or are not in their library and to reflect the operational consequences of these declarations. The mechanisms for doing this need to be applicable for all levels of users, from individuals and small groups, up to the Library of Congress. There will always be a need for information users to assess the validity and trustworthiness of materials, and these should be provided as part of the suite of library services, rather than as an entry condition for material being in the library.

5.4. Individual information items will be provided, stored, processed, and provided in multiple representations.

Our mechanisms can not assume canonical forms of information sources without unduly restricting the materials that can be accessed. Information in the digital library is inevitably going to be heterogeneous and we need mechanisms that deal with derivations of meta-information and cross-translations.

5.5. Materials in the digital library will be in a variety of media forms, including text, image, audio, video, 3-D models, etc.

Since we are doing the "glue" rather than a comprehensive application system, we will not be providing extensive facilities for each of these. Our underlying model must take into account their diversity (e.g., dealing with time-sequence media and multiple resolution media) in developing appropriate forms for meta-information and protocols for access and services.

5.6. Meta-data is content.

A significant part of the information in a digital library is the meta-information about items (both on-line and physical), such as annotations, catalogs, indices, descriptions of indices, etc. We should consider this kind of information as a "first-class citizen" which is managed in the same way as other information (e.g., access control, archiving, source attribution, etc.)

6. PLATFORMS

6.1. In the long term, the digital library will be based on platforms that are standard in the commercial world. However, in the course of this project we will best enhance the research by using platforms that the people in the project have familiarity and skills in using.

At the moment it appears that the long-term universal platform is likely to be based on Microsoft Windows and its descendants. Product development for long-term use needs to take this into account and focus on the paths that will be facilitated by the growth of the major markets. But as stated in our initial
section, we are building a research vehicle, not a long-term product. By using unix (for servers at least) we leverage our existing base of knowledge and skills.

6.2. We are designing for an audience that has substantial communication bandwidth (e.g., sufficient to run Mosaic-like browsers).

Although most of the world does not have this level of connectivity and may not for quite a while, our testbed community does. We believe that this level will be the minimum that people will have within a few years (if they are connected at all), and that we should develop with that target in mind, rather than being concerned with compatibility with minimal (under 10KB, non-graphic) connections.

6.3. A key issue in the overall design is a concern with compatibility and interoperation with other standards and integrative systems.

In a project of this scope we are not going to develop a competitor to major vendors or development projects. This means that either we have to do a limited enough experiment that we can pull the pieces together ourselves, or do it in a way that can take advantage of existing systems that in some partial way provide things we are looking for. We choose the latter, as long as the desire for interconnection doesn't get in the way of the basic research goals.

Change history:
Minor corrections by Terry Winograd, June 24, 1995.
Original version February 23, 1995 by Vicky Reich and Terry Winograd