From Informal Resource Discovery to an Information Infrastructure
Michael F. Schwartz
University of Colorado - Boulder
To be Published in Internet Society News 2(3), Fall 1993

The U.S. National Information Infrastructure (NII) Act of 1993 calls for broad deployment of information applications in education, health care, libraries, and government data. This represents an enormous call to action, requiring development of the applications software and the physical network, as well as efforts to bring historically isolated agencies and information pools onto the Internet.

Where do current Internet-based information systems fit into this picture? Systems like Gopher, WAIS, and WWW have gained loyal followings, and are widely heralded as important new Internet applications. They have caught the attention and imagination of politicians and the media, and offer an immediate means of bringing some of the data targeted by the NII online. However, significant advances are needed both in the information systems and in the underlying communications substrate before a true "information infrastructure" can exist. I'll start with the systems problems.

First, the information provided by current resource discovery systems is mostly informal, with no guarantees about data validity or timeliness. This is not surprising, given the "grass-roots" manner in which most of the data are placed in these systems. To some extent, this problem can be corrected by formalizing (and funding) a support function. But there are technical problems as well. Most of the data have no associated attributes recording age, seal of authenticity, or even record of origin. The systems simply lack the ability to record or act on this information at present.

Second, current systems provide little or no information structure that could support a wide range of automated data processing tasks. The vast majority of information provided by these systems is intended for human consumption - formatted text or typeset graphics. Often the data were derived from more structured sources (such as a relational database or spreadsheet), but the disseminated data lack structure. More importantly, the current generation of Internet tools provide no means of representing or accessing this structure. Mostly they provide flat keyword searching and menu-browsing interfaces.

Third, the current systems will not scale well for use in a markedly expanded Internet. Already it is difficult to navigate around a large Gopher menu, or choose appropriate keywords when searching a large WAIS collection. Imagine what will happen when there is one million times as much information available. Other scalability problems exist as well, most notably in the demand placed on servers of widely popular data.

A number of research and development groups are working on these systems problems, in both the Internet Engineering Task Force and the Internet Research Task Force. The challenge will be to address these needs while retaining the practical ease of installation and use of the current Internet discovery tools. In particular, while X.500 addresses many of these problems, at present it still requires a good deal of administrative and computational resources to install an X.500 server. X.500 servers are deployed at a much slower rate than the other discovery services.

In addition to the systems problems, work must also be done on the lower network layers. Perhaps most importantly, we need support for commerce-grade (authenticated and private) communications. Until this is routinely available, no hospital could possibly risk sending patient medical history data across the network. The communications substrate will also need to support quality of service guarantees, so that, for example, a medical image being transmitted across the network will not be delayed by a less important background file transfer. Of course, an important component of providing such guarantees is upgrading the available network bandwidth to meet the increased data demands.

One final problem area is the lack of a workable business model for developing and deploying information services. There has been a history of commercial development and deployment difficulties to date. For example, witness the poor responses to "videotext" offerings from the U.S. regional telephone companies, and to attempts to commercialize the development of current generation Internet information tools. With the beginnings of Internet-based media service offerings (e.g., the creation of Electronic Newsstand Inc. over Summer 1993), a business model will hopefully arise that can help lead the way to successful commercial development of the NII.