□Multiple Standards? No problem.

(A position paper for Digital Libraries '94)

Mitchell N. Charity

Library2000 Group, Laboratory for Computer Science, Massachusetts Institute of Technology, 545 Technology Square, Cambridge MA 02139 mcharity@lcs.mit.edu[1]

Abstract

It is neither necessary, nor desirable, for the developing global information infrastructure to be built upon a small number of carefully crafted standards and protocols. The developing infrastructure is well suited to incremental evolution, with an interoperating multiplicity of standards and protocols, and the associated benefits of heterogeneity. However, this does require that intentional isolation be uncommon.

Discussion

As we develop a global information infrastructure, we look to standards to contain the costs of interoperation. There are multiple approaches to standardization, each with benefits and disadvantages. As this is a time of beginning, many argue for "getting it right to begin with", for rapidly standardizing on a small set of well thought-out protocols and practices. I suggest this is the wrong objective. The history of the Internet illustrates the advantages of standardization based on rough consensus and working code. Further, I suggest having multiple and imperfect standards should be treated as a design goal. The global information infrastructure is particularly well suited for, and indeed I suspect, will almost unavoidably be characterized by, incremental evolution with a multiplicity of protocols, all interoperating using gateways, multi-lingual participants, and third-party value-added services.

The standardization of the Internet illustrates the range of approaches to standardization available to us. The Internet is built on standards organized through the IETF, the Internet Engineering Task Force. It is an open, informal organization, which creates standards based on rough consensus and working code (some say working consensus and rough code). The IETF standards are advisory, and sometimes the marketplace does not adopt them. Similarly, the IETF may explicitly leave a choice to the marketplace when there is no clear technical basis for choosing between alternatives. This is in contrast to the International Standards Organization approach of highly structured communication, used to generate specifications, followed by implementation attempts and required compliance. The IETF model, while arguably weak in envisioning big steps, admirably coordinates evolutionary change in an environment where many solutions are being attempted.

I suggest a vision of a global information infrastructure which is an inhomogeneous, but interoperating, network of competing standards and services. In the face of heterogeneity, interoperability is maintained by gateways/bridges, widespread multi-linguality, and third-party value-added services. Gateways take information available via one protocol and make it available via another. Clients accept, and servers provide, information via multiple protocols. Third-parties take information and add to it, and provide glue services like namespaces and notary validation.

Consider the recent history of the Internet. A variety of standards and protocols are in use - WWW (HTML and HTTP), Gopher, WAIS (Z39.50), FTP, Email, News, and others. They provide overlapping functionalities. They can be combined in various ways. There are gateways between them. Clients usually, and servers often, understand several of them. As one or another provides better service, clients and resources migrate across gateways to become native and utilize the new functionality. Communities of interest extend and create standards to satisfy "local" needs. These mutations then either die out, through replacement or old age, are adopted into the standard, or are forever gatewayed, as when they are worth the cost of persistence, but not of adoption. But why does this heterogeneity exist?

A number of factors have contributed to a state of integrated heterogeneity, and also reveal limits to this approach. The two common threads are easier system creation, due to a growing programmer base and richer software tools, and easier system propagation, due to the spread of computers and networks. For instance, consider the creation of a new gateway. Creation, rather than operation, is generally the obstacle. But all it takes is the commitment of one person, somewhere in the world, to

[1] The Library2000 Group, under Professor Jerome Saltzer, is working on the system engineering of global-scale digital libraries. This work was supported in part by the Corporation for National Research Initiatives, using funds from the Electronic Libraries project of the Advanced Research Projects Agency of the United States Department of Defense, under CNRI's grant MDA972-92-J1029.

build a bridge. The net then permits universal access, largely independent of geography, and facilitates the replication of the service. Standards are merely a way of reducing the number of such people and bridges required.

But there are limits. Some lack of interoperability is intentional, isolation conferring some competitive advantage to some or all parties in the isolated group. Here standardization can play a cartel breaking role, reducing the cost of escape, and facilitating diffusion by lowering the obstacles in the technical background against which political decisions are made. Integrated heterogeneity requires bridges, and bridges require that data providers be willing, and technically/legally able, to permit third-party reselling. Finally, it may not be possible to gateway very complex (i.e., badly designed) protocols.

In summary, I suggest a variety of competing standards and protocols are a sign of a healthy global information infrastructure. Difficulties stem instead from standard fragmentation and dialectic incompatibilities, and from architecting standards and practices without emphasizing interoperation.