INFRASTRUCTURES FOR SHARING GEOGRAPHIC INFORMATION AMONG ENVIRONMENTAL AGENCIES

by

John D. Evans

Submitted to the Department of Urban Studies and Planning
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Information Systems in Planning

at the

Massachusetts Institute of Technology

June 1997

Signature of Author ____________________________________
John D. Evans
May 30, 1997

Certified by ____________________________________
Joseph Ferreira, Jr.
Professor of Urban Planning and Operations Research

Accepted by ____________________________________
Frank Levy
Chair, Departmental Ph.D. Program

Infrastructures for sharing geographic information among environmental agencies

by

John D. Evans

Submitted to the Department of Urban Studies and Planning
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Information Systems in Planning

ABSTRACT

This research draws on several organizational and technological perspectives to examine the design and growth of infrastructures for inter-organizational geographic information sharing, and their role in collaborative environmental management. The study draws, first, on the experiences of selected coalitions of government agencies to discern their organizational dynamics; and on a software prototype that illustrates coming technological trends. The study gives special attention to geographic information, which often requires special handling and promises particularly important influences on organizations and joint policy-making. It also seeks to understand the interdependence between human and technical aspects of geographic information infrastructures.

The first research phase is a case study of three groups of government agencies that are building networked information-sharing systems for the joint protection of large ecosystems (the Great Lakes, Gulf of Maine, and Pacific Northwest rivers). These cases richly illustrate the challenges and benefits of designing flexible standards, rethinking organizational structures, and adjusting decision-making processes to depend on shared geographic information. The study's second phase, a prototype networked service for digital orthophotos, suggests that shifting the focus of information sharing from datasets to data services is becoming increasingly feasible, but that organizations may need to adapt to new forms of information sharing. Together, these findings suggest an expanded view of standards as layered, strategic choices; and of organizations in complex, interdependent relationships.

Thesis Supervisor: Joseph Ferreira, Jr.
Title: Professor of Urban Planning and Operations Research

Table of Contents

Tables and Figures

Acknowledgments

"Begin at the beginning," the King said gravely,
"and go on till you come to the end: then stop."
- L. Carroll, *Alice's Adventures in Wonderland*

I'd like to thank my advisor, Prof. Joseph Ferreira, Jr., for sticking it out with me all these years and giving liberally of his time, especially in the home stretch. My thanks also go to Prof. Wanda Orlikowski, of MIT's Sloan School of Management, for opening up new ways to do research, and for restoring my hope with three simple words: "important, fascinating work!" And I thank Prof. Lyna Wiggins, at Rutgers University, whose straight talk got me back in motion several times when I was fogged in. Special thanks to fellow denizens of the Computer Resource Laboratory, especially Lijian Chen, Kamal Azar, and Annie Kinsella, who brightened my graduate school experience with their humor and friendship. Thanks to my wife Suvia for standing with me through a lot more than she'd bargained for. And thanks to God, "our dwelling place through all generations."

This material is based upon work supported by the National Science Foundation, under Grant No. SBR-9507271, and by the 1995 Competitive Cooperative Agreements Program of the Federal Geographic Data Committee. Any opinions, findings, conclusions, or recommendations expressed in this material are my own and don't necessarily reflect the views of the National Science Foundation or of the Federal Geographic Data Committee.

Chapter 1: Introduction

What does it take to share geographic information? Such sharing would seem especially fruitful in the arena of environmental protection and policy, where key decision variables are often linked by physical pathways (waterways, land-forms, habitat) that traverse jurisdictions, industries, hierarchies, and other territorial lines. Geographic information would seem especially conducive to sharing, due to the high cost of producing it, its potential for widespread re-use, its value in spatial-analytical overlays, and its often unique role in organizational structures. Also, the coming-of-age of the Internet and increasingly sophisticated network software ought to make geographic information sharing especially easy, convenient, and powerful.

Yet it's unusual to see planners or public managers meaningfully sharing information across organizational boundaries. The reasons for this would seem to be part technical, and part organizational, and often peculiar to the nature of geographic information: complex in structure and interpretation, rich in meaningful inter-relationships, and difficult to understand or use without special-purpose tools.

An important solution to this dilemma may be a geographic information infrastructure—that is, an ongoing, multi-purpose mechanism created to help members of participating organizations make use of each other's geographic information. Learning how to design and grow such infrastructures, within a rapidly evolving technological and organizational context, seems key to effective inter-agency collaboration and sharing of geographic information. This dissertation examines the design and growth of organizational and technological infrastructures for sharing geographic information. To remain grounded in real organizational experience, yet sensitive to rapidly changing technology, the research takes a hybrid approach, which draws both on social / behavioral perspectives and on GIS technology, standards, and networking.

Chapter 2 presents these various perspectives as they relate to the question of building and sustaining geographic information infrastructures. Chapter 3 explains the methods I followed for my "hybrid" study comprised of organizational case studies and software prototype development. Chapters 4, 5, and 6 then describe three case studies of existing inter-organizational infrastructures for sharing geographic and other information. Chapter 7 follows up with a synthesis that analyzes findings across the three cases in several passes, for an understanding at several different levels. Chapter 8 takes a very different angle on the same topic: it reports on a year-long effort to create a networked service for digital orthophotos, and traces the implications of that effort for the design and growth of geographic information infrastructures. Finally, Chapter 9 draws on the preceding organizational and technological findings to sketch more general implications for technology, organizations, and policy, and directions for future research.

Chapter 2: Background and Scope

1. Overview

A priori, the problem of geographic information sharing would seem to be part technical, part organizational, and often tied to the nature of the information and its use. The following paragraphs present definitions of information sharing and infrastructures, and trace several related lines of research, from sociological variance and process theories to interoperability, standards, and the coming-of-age of Internet tools and services. These are tied together by a research perspective based on structuration, which helps to interpret the case study findings and prototype results and to draw useful lessons for technological and organizational infrastructure design.

2. Concept Definitions

What is meant by information sharing, anyway? And what role do infrastructures play? Before reviewing related theories and research, it helps to clarify the meaning of these two key terms.

a. *Various conceptions of sharing*

"Sharing" can mean a great many things to people. In principle at least, everyone is in favor of it; it's a lesson we learn in kindergarten: "Share everything. Play fair. Don't hit people" (Fulghum, 1987). Roget's Thesaurus (Longman, 1994) associates the word "shared" with activities which are in common, cooperative, or participatory, and with entities that are merged, composite, coincident, in partnership, or "in the same boat." When it comes to sharing *information*, related words include "notification," "exchange," "pooling," and "dissemination." Carter (1992) contrasts two kinds of information sharing in particular: the buying and selling of information products, and the use of multipurpose information by members of a partnership. Tosta (1992) emphasizes non-commercial "vertical" relationships (as between local, state, and federal government) and "horizontal" relationships (linking organizational "peers" such as adjacent cities). Frank (1992) takes a more conceptual view, suggesting that "the specific problem of sharing data is [...] to communicate meaning in a situation where the perceived reality differs because the two organizations have different goals and work on different tasks."

This research examines information sharing among organizations with overlapping geographic areas of interest. This is not really "vertical" sharing because the participants don't necessarily include, constitute, or report to each other; some are even in different countries. Neither is it "horizontal" sharing: participants span a variety of levels, from towns and schools to state offices, regional associations, and federal agencies. The sharing relationship generally links groups that have different goals and tasks, but whose concern for a valuable shared natural resource encourages them to work together, to coordinate their efforts, to learn from each other, and to build some kind of infrastructure to facilitate these activities.

b. Information sharing infrastructures

Most often, information sharing occurs in an *ad hoc* fashion, via methods that are devised anew with each interchange: that is, single-use, single-purpose mechanisms. However, as the volume and frequency of information sharing grows, a more permanent mechanism often becomes beneficial, one that can function repeatedly and serve a variety of purposes. This multi-use, multi-purpose mechanism is what I have termed an *information sharing infrastructure.* Some use the word "infrastructure" in a specific technical sense, to mean "the basic facilities, services, and installations needed for the functioning of a community or society, such as transportation and communications systems, water and power lines, and public institutions including schools, post offices, and prisons" (Houghton-Mifflin, 1992). Others define it quite broadly, as "the underlying foundation or basic framework of a system or organization" (Merriam-Webster, 1990). The U.S. National Telecommunications Infrastructure Administration uses the following definition:

> The telecommunication networks, computers, other end-user devices, software, standards, and skills that collectively enable people to connect to each other and to a vast array of services and information resources" (NTIA, 1996).

The special case of geographic information is of particular interest in sharing and infrastructure-building, due to the high cost of producing it, its potential for widespread re-use, and its value in spatial-analytical overlays (Evans and Ferreira, 1995). In particular, since 1990, the U.S. Federal Geographic Data Committee has devised and promoted a National Spatial Data Infrastructure (NSDI) defined as

> (...) (1) standards to facilitate data collection, documentation, access, and transfer; (2) a basic framework of digital geospatial data that meets the minimum needs of large numbers of data users over any given geographic area; (3) a clearinghouse to serve, search, query, find, access, and use geospatial data; and (4) education and training in the collection, management, and use of geospatial data.
> (Tosta, 1994)

This definition, specific to geographic information, includes several elements not mentioned previously: in particular, it adds a "framework" of generic digital data, built to serve a wide variety of purposes — similar to, though less specific than the "multipurpose cadastre" called for in the early 1980s (National Research Council, 1983). According to proponents of the National Spatial Data Infrastructure, building it will depend on significant organizational changes — in particular, formal partnerships to divide responsibilities, costs, benefits, and control among several organizations (Tosta, 1994; National Research Council Mapping Science Committee, 1994).

The information sharing infrastructures I refer to in this study roughly follow the NSDI definition above, though at a regional, rather than a national, scale. An information-sharing infrastructure, as described here, links organizations with separate goals and tasks by means of standards, navigation and conversion tools, shared "framework" information, and institutional structures such as partnerships. A central question of my research is how best to design, build, maintain, and grow such infrastructures for effective geographic information sharing in support of environmental policy, planning, and management.

3. Related research

Several areas of recent research provide guidance for building and maintaining infrastructures for geographic information sharing. The next few paragraphs outline organizational and technological aspects of geographic information sharing, which together provide a useful language to describe and compare the design choices and development styles adopted in the various cases, and to formulate useful prescriptions for the future. The review of these research areas is followed by a research perspective which emphasizes their reciprocal influence on each other.

a. Organizational aspects of geographic information sharing

Organizational aspects of information sharing can be drawn from the large body of literature on collaboration, consensus building, and coordination. Much of this literature emphasizes a *variance* model which assumes that changing the levels of certain input factors leads, by some functional relationship, to variations in performance outcomes. The following paragraphs contrast this approach with a *process* model, which traces micro-level decisions over time to explain how observed outcomes come about.

i. Factors in information sharing, consensus, and coordination.

In the view of many practitioners, the greatest obstacles to information sharing seem to be behavioral, rather than technical (Croswell, 1989). The literature (reviewed by Grandori and Soda (1995), Pinto et al. (1993), Mizruchi and Galaskiewicz (1993), Harrigan and Newman (1990)) shows a complex set of factors affecting inter-organizational collaboration: some of these factors pertain to individuals, some to organizations as a whole, others to relationships between organizations, and still others to the broader political and social context (Schermerhorn, 1975). The benefits of collaboration for individual organizations include economies of scale, lower overhead, reduced risks — all in the interest of increasing efficiency and thus surviving in a complex ecology. Collaboration is also facilitated by a clear interdependence between two or more organizations: this interdependence may involve pooled resources, sequential tasks, or resource transfers (Alexander, 1995).

The picture is no less complex within the specific area of geographic information sharing. For instance, Rhind (1992) has argued that effective information sharing depends on its appropriate pricing — in contrast with Epstein (1995), who states that free dissemination is what's needed. At the level of inter-organizational relationships, information sharing may be facilitated by negotiation (Obermeyer, 1995), a common, "super-ordinate" goal (Pinto and Onsrud, 1995), a "killer application" (Brodie, 1993), and clear data ownership (Carter, 1992). Technical abilities in an organization can make a big difference, for example when organizations make quick technical choices to support limited internal purposes (Craig, 1995) or rely on inadequately trained staff (August, 1991).

ii. Processes of information sharing, consensus, and coordination.

In contrast with the variance approach, Van de Ven and Walker (1984) and others suggest that a process approach, which traces micro-level decisions over time, is better than a macro-level predictive model at understanding patterns in a complex inter-organizational setting. One way of reconciling the macro and micro levels is social network theory (Granovetter, 1972), which compares the influence of strong and weak social relationships in transmitting information. Boland and Tenkasi (1995), in their cognitive study of organizational communication, link shared information to consensus-building through *perspective-making* (articulating one community's knowledge) and *perspective-taking* (understanding the knowledge of another community). To facilitate inter-group communication, they suggest the use of "boundary objects" (maps, structured narratives, and the like) to help structure one's world-view and relate it to another's. In a similar vein, Gray (1985) separates the formation of collaborative relationships into a three stage process — problem-setting, direction-setting, and structuring — each with its own conditions for success.

b. Technological aspects of geographic information sharing

Technological aspects of information sharing are the subject of a diverse array of recent research in multi-database theory, standards, and distributed data networks.

i. Levels of connectivity: physical, logical, semantic

In general, data storage and retrieval systems may be linked by three types, or "levels," of connectivity (Wang and Madnick, 1989). Data networks provide *physical* connectivity — basic communication between computers. Sharing structured information between computer systems requires *logical* connectivity—the ability to reconcile data models (Batini et al., 1986; Peckham and Maryanski, 1988) and query procedures (Litwin et al., 1990) between database systems. Finally, *semantic* connectivity (Siegel and Madnick, 1991) bridges differences in data definitions, for accurate interpretation and use of the information itself. Interoperable database systems, offering connectivity at all three levels, have been the focus of intense research (Litwin et al., 1990; Sheth and Larson, 1990; Templeton et al., 1987; Hsu et al., 1991). Interoperability is defined in a variety of ways, but it refers in general to the ability of different software systems to understand each other's information requests and respond to them appropriately. Logical and semantic connectivity are the emphasis of data repositories (Jones, 1992), which extend data dictionaries (Narayan, 1988) through the use of enterprise models (Sen and Kirschberg, 1987). Most work on interoperability has been with alphanumeric data; but some geographic extensions to data dictionaries (Marble, 1991) and repositories (Robinson and Sani, 1993) do exist, although the semantics of geographic data remain an open research question, both theoretical (Firns, 1992; Nyerges, 1991; Morehouse, 1990) and practical (Baker and Broadhead, 1992). Geographic information, with its multidimensional data structures and relationships, may present quite different interoperability issues from its alphanumeric counterpart; yet multi-database concepts and methods are useful for understanding and comparing geographic information sharing systems (Nyerges, 1989; Mackay and Robinson, 1992). In the area of geographic information, the OpenGIS Consortium (Ganter, 1995) has proposed interoperability through a generic set of geographic data entities and geographic data processing actions. Thus, any two software systems can interoperate once they've defined their own data entities and manipulations in terms of the generic ones. Data queries and their replies can then pass through a "broker" that interprets them for use by their recipients. Even though a global model of geographic data and analysis is still far from complete, brokers based on generic, object-oriented models are a promising approach to bringing together autonomous data services. Thus, even tentative, partial steps in this direction can have significant payoffs.

ii. Standards, metadata, and non-intrusive sharing

Data compatibility issues (e.g. differing formats, definitions, or scales) often complicate the task of bringing together separate information resources for comparison, summary, or analysis. Traditional standards (U.S. National Mapping Standards, Spatial Data Transfer Standard) are a valuable *lingua franca* for sharing geographic information; but sharing is often needed between autonomous organizations with different, yet well established procedures or quality requirements. This need is addressed by newer, flexible standards based on *metadata* (that is, structured data descriptions) (Evans et al., 1992) and *queries* (i.e., structured data requests) (Egenhofer, 1992). These provide a shared vocabulary, a "minimal standard" that defines the user's interaction with information without intruding on the information itself. Noteworthy efforts include the US Federal Content Standard for Geospatial Metadata (FGDC, 1994), the Common Object Resource Broker Architecture (CORBA) (Siegel, 1996), and the Open Geodata Interoperability Specification (OpenGIS) (Ganter, 1995). Although little is known about the tradeoffs of these minimal standards vis-à-vis more intrusive methods, they could redefine the whole notion of data sharing.

iii. Network tools and network management

As the Internet has come of age, so have tools for locating, retrieving, filtering, and using networked information, from the file transfer protocol (*ftp*), network file systems, and archive catalogs (*archie*) (Emtage and Deutsch, 1992), to Wide Area Information Servers (Kahle, 1991), Gopher (Schwartz et al., 1992), and especially the World Wide Web (Berners-Lee et al., 1992). Several tools have been developed specifically for locating and viewing networked geographic data (Walker et al., 1992; Menke et al., 1991). Tools of this kind underlie several information sharing infrastructures in existence today, and continue to promise scalable, flexible ways to distribute geographic data across a growing network. Frank (1994) sketches the role of data catalogs and navigational tools in several possible scenarios of a future spatial data infrastructure.

c. Interdependence of organizational and technological aspects

Sharing geographic information is often characterized as either a technological problem or an organizational one. For example, an agency may not make use of an outside data source for lack of suitable data-conversion software or a means to search through reams of unknown data; but information sharing is often seen as impeded by "turf battles," institutional inertia, or unclear intellectual property. In response, most researchers have focused on either the technical or the organizational topics described above. These focused research efforts are valuable in their own right; but in an unsettled, rapidly changing technological and organizational context, sharing geographic information is rarely a purely technical problem or a purely organizational one (Evans and Ferreira, 1995). For instance, the apparent "inertia" that hinders the use of outside geographic data may in fact be a quite sensible response to difficult data-coordination problems intimately tied to current technology and to the complexity of the data itself. Conversely, technical innovations such as search engines or data-conversion interfaces may only affect information sharing in an organization that encourages its members to explore new approaches to their work.

The use of geographic information can have its own peculiar influence on organizations, for two reasons. First, as Goodchild (1992) articulates, "what distinguishes spatial data is the fact that the spatial key is based on two continuous dimensions." Consequently, using geographic information usually requires interpolating values between known data points: a matter of interpretation and analysis, quite different from simply looking up alphanumeric information in relational data tables. Second, Chorley (1988) and others note that geographic information is most useful when it is linked or merged with other geographic information, and that most geographic information is used for more than one purpose. This makes it hard to anticipate and codify how geographic data are retrieved, combined, and used, and it blurs the distinction between information managers, analysts, and users. This would tend to contradict most hierarchical, departmental organizational structures, and encourage more fluid, overlapping task-oriented structures.

In describing the relation between technologies and organizations, Markus and Robey (1988) emphasize an *emergent* perspective, focused on the interaction between organizations and technology, in contrast with both a *technological determinism* (in which technologies are presumed to have known, inexorable effects on organizations) or a social *strategic choice* (in which technologies are seen as inexorably shaped by organizational intentions and actions). Barley (1986) invokes *structuration theory* (Giddens, 1984) to trace the ongoing, recursive influence between an organization's rules and resources and the behavior of its members, as triggered by new technologies. Using the structuration perspective, DeSanctis and Poole (1994) emphasize the "intertwined" nature of technological and behavioral patterns, and Orlikowski (1992) proposes a useful view of technology as a malleable structural property of organizations. That is, within a structuration perspective, organizational intentions alone cannot give rise to a given technology, nor can a technology have a fully predictable effect on organizations. Rather, in every phase of a technology's existence — its conception, design, deployment, use, evaluation, and modification — the human actors involved mediate both causal effects in unpredictable ways.

This perspective seems like a useful one for the study of complex phenomena such as inter-organizational information sharing, collaboration, or consensus: it enables the study of technological and organizational aspects in concert, rather than in isolation from each other. Accordingly, this research examines both technological and organizational design choices and contexts; their mutual interaction; and their joint relationship to planning and policy. Furthermore, within this perspective, the technical design of an information sharing infrastructure cannot be evaluated apart from its implementation and use within a group of organizations. Thus, any solutions considered here should be conceived as *packages* of mutually-influencing technological and organizational features, and as *pathways* of not-fully-predictable growth and change over time.

d. Organizational and technological change

Indeed, not only do technologies and organizations have a mutual, recursive influence on each other, but they both tend to undergo continual or periodic change. Yet designers of information systems often make the more-or-less tacit assumption that organizations are static — that structural changes are abnormal and reach an equilibrium. Conversely, organizational thinking often tends to accept technologies as artifacts with a fixed role and stable features. However, particularly as seen through the structuration perspective described previously, social structures undergo constant change, and information technology itself is an element of that social structure, enabling some actions, constraining others, and itself shaped by those very actions over time (Orlikowski, 1992). Particularly in the case of large information networks, this perspective implies "organic, yet systematic" change over time, as Spackman (1990) notes in his discussion of networked organizations:

> Our road and rail networks and the international telephone network have grown organically, yet systematically, over many, many years. The integrated corporate information network is just such an enterprise and, in one sense, it already exists, while in another, it will never be complete. The issue therefore is not how to *build* such a network, but rather how it is allowed to *grow*.

In this research, both technological and organizational change are considered normal and are assumed to be ongoing. This dynamic perspective has two important implications. First, the technical and organizational design of information sharing infrastructures are less a set of fixed, interlocking components than a chosen *direction*, or even a *style* of evolution through an uncertain future. Second, although a broad set of "levers" can be pulled to affect sharing, collaboration, or consensus, their influence on outcomes is uncertain and only temporary.

Chapter 3: Research Methods

1. Overview

Based on the technological and sociological perspectives presented in Chapter 2, I embarked on two research activities: *(i)* observing the deployment and use of current technologies in an intergovernmental context, and *(ii)* experimenting with new and emerging technologies in a laboratory setting. This hybrid approach recognized the intertwined nature of technologies and organizations in geographic information infrastructures. Furthermore, by observing organizations in the recent past and building tools in the laboratory for the near future, I sought to build an understanding that would hold across a range of contexts and remain valid for some time to come.

a. Case studies: contextual views of information sharing infrastructures

In the first phase of the research, I studied three interagency infrastructures for information sharing related to large shared natural resources: the Great Lakes, the Gulf of Maine, and the Columbia River system. The case study traces the design choices people made in building these infrastructures, and the infrastructures' patterns of growth and change, so as to understand their role in information sharing, collaboration, and consensus-building among the organizations involved. Section 2 below details my choice of cases and the methods I used for studying and learning from them.

b. Prototype development: anticipating the effects of emerging technologies

The second phase of the research drew on a prototype networked information service I built as part of the US National Spatial Data Infrastructure. This service provides detailed geographic data and metadata to users via the World Wide Web, and embodies several design choices likely to facilitate high-quality information sharing. As such, it provided a tangible response to some of the needs arising from the case study, and it suggested the trends likely to bring radical changes in the meaning and modes of information sharing in years ahead. Section 3 below sketches the prototype design and development effort.

2. Organizational case studies

The first phase of research was a study of three inter-governmental geographic information infrastructures. I sought to *compare* the design, implementation, and evolution of these information sharing infrastructures, *understand* their patterns of structural change over time, and *document* any specific new tasks or processes enabled by information sharing. I also wanted to focus on cases that could serve as examples for a wide range of organizations, and where geographic information sharing seemed likely to grow in its impact on the tasks and processes of participating agencies. These goals, and the need to study a complex topic within its institutional context, led me to adopt a case study strategy (Yin, 1984). The next few paragraphs sketch the criteria I used for evaluation and comparison, my choice of cases, data collection methods, and methods for analysis and synthesis.

a. Criteria for evaluation and comparison

I chose four criteria to evaluate and compare geographic information infrastructures:

1. The *impacts* of the infrastructure, in terms of changed *tasks* (jobs performed that would have been infeasible without the infrastructure), or changed *processes* (jobs performed differently thanks to the infrastructure).

2. The *size* of the infrastructure, based on its data holdings, traffic volumes, budgets, etc.;

3. The *quality* of the *shared information*, in terms of concurrency and timeliness, its precision and accuracy, and its encapsulation (details below).

4. The *quality* of the *information-sharing mechanism*, in terms of reciprocity, scalability, and non-intrusiveness.

These last two criteria require some clarification. First, the quality of shared information for a particular use is related to its concurrency, timeliness, and encapsulation:

- *Concurrency and timeliness:* how well does the update frequency of the shared information match the change frequency of the phenomenon it represents, or that required by users for their tasks? For instance, given that weather patterns change every few hours, a newspaper's weather report has a lower concurrency than National Weather Service broadcast. Timeliness depends on the user: a mariner might listen to the Weather Channel continuously to know when to return to shore, but a climate researcher may only be interested in yearly summaries.

- *Precision and accuracy*: how exactly specified is the information (decimal digits, spatial resolution) and how correctly does it describe phenomena (percent error, reliability)? For instance, my digital watch tells time with much more precision than my antique grandfather clock, but if I set it wrong, it may give me far less accuracy.

- *Encapsulation*: how much effort is needed to use the shared information once it's obtained? Not too long ago, the US Census Bureau distributed its data as strings of EBCDIC numbers on 9-track magnetic tapes; any use required lengthy parsing and repackaging. It now puts out much more encapsulated data: its Web sites let anyone query and retrieve Census data in a variety of formats, or even create demographic maps on demand.

These three criteria form a user view of the quality of information sharing. A second perspective is that of the system's architect or manager, and assesses the quality of the sharing mechanism itself, which depends on its reciprocity, scalability, and non-intrusiveness:

- *Reciprocity:* how readily can participants both send and receive information? For instance, with a newspaper or television broadcast a source sends information to receivers; receivers have few opportunities to communicate back to the source, or to other receivers. By comparison, a book club or online forum is more reciprocal in that it allows its participants both to give and receive information.

- *Scalability:* how much maintenance effort is needed to add large numbers of participants or datasets, or wide varieties of data and protocols, to the infrastructure? Some databases are quite rigidly specified: each additional user or data source requires as much work to set up as the first user or source. More scalable systems can grow in size or use with much less incremental effort.

- *Non-intrusiveness:* how little change does information sharing impose on participants' internal work procedures? Accessing a database, and including its information in an ongoing project, might require very specific formats, software, or even operating systems or hardware: such an infrastructure would be quite intrusive. Recent "open systems" and "interoperability" efforts have sought to facilitate the non-intrusive transfer of information and instructions between independent computer systems.

b. Choice of cases

As mentioned earlier, I wanted to focus on cases that could serve as examples for a wide range of organizations, and where geographic information sharing seemed likely to grow in its impact on participating agencies. Therefore, I chose sites for study according to the high quality of their information-sharing, as outlined above, and I applied several additional selection criteria. First, to focus on task-related information sharing, I favored *non-supplier* organizations, ones whose purpose was research, policy, or management, over ones that existed simply to provide data to others. Second, I favored cases where an *infrastructure* of some kind was being built to enable systematic, planned information sharing (i.e. not just *ad hoc* sharing, via methods invented anew each time). Third, I favored *loosely-coupled coalitions* over strong "umbrella" agencies that could decree uniform standards and procedures. Fourth, I chose agency groups that had formed around a clearly identified *shared natural resource:* this provided a "superordinate goal" that would minimize organizational resistance to change. And fifth, I favored cases where *geographic* information was being shared. So in summary, I looked for loosely-coupled coalitions of non-supplier agencies that were building an infrastructure to share geographic information in managing a shared natural resource.

To find suitable cases for study, I obtained initial contacts around the country from colleagues and a few Boston-area Federal offices. These people in turn referred me to their colleagues and acquaintances who knew of geographic information sharing efforts in environmental contexts; and after a few cycles of this, I learned of several interagency information sharing efforts motivated by shared ecosystems. I spoke by telephone with the information systems or GIS coordinators for the Chesapeake Bay, Gulf of Mexico, and Great Lakes programs of the Environmental Protection Agency (EPA); the Gulf of Maine Council on the Marine Environment; the Northern Forestlands

Study in northern New England; the Silvio O. Conte Wildlife Refuge on the Connecticut River; the Delaware River National Estuary Program; the Bonneville Power Administration's Northwest Environmental Database; the Lake Tahoe Regional Planning Agency; and the Tennessee Valley Authority's TERRA decision support system.

Two findings quickly became clear from this preliminary inquiry. First, in order to have *any* cases to study, I needed to relax my "quality" criteria, which were quite stringent and applied to a more highly networked context than that of these agencies.[1] Second, several of these programs were not building an infrastructure: their information sharing consisted of either *ad hoc* exchange of files on diskettes, a one-time effort to compile information from many sources, or a homogeneous distributed information system within a single agency. Thus, I narrowed my field to the Great Lakes Information Network (GLIN), the Gulf of Maine Environmental Data and Information Management System (EDIMS), and the Northwest Environmental Database (along with its successors, the Coordinated Information System and StreamNet).[2] These were enough for the exploratory analysis sketched in *d.* below. A brief sketch of the three cases is helpful before I describe my chosen research methods.

- *Great Lakes Information Network (GLIN).* The Great Lakes Commission, a partnership of eight US states, began experimenting with the Internet in 1993 to enhance communication and coordination among the many groups concerned with the Great Lakes. The resulting Great Lakes Information Network (GLIN) (cf. Ratza, 1996) links the Great Lakes states (Minnesota, Wisconsin, Illinois, Indiana, Michigan, Ohio, Pennsylvania, and New York), the Canadian province of Ontario, and several Federal agencies and other public and private groups in both the US and Canada. GLIN saw rapid, unpredictable growth in size and usage through the "evangelistic" efforts of its founders and the nature of the emerging World Wide Web, and encountered challenges related to its growth, including distributed data management and interaction with other information sharing initiatives.

---

[1] This was mid-1994: the Web was already growing rapidly among universities and research centers, but most government agencies were not yet online.

- *Gulf of Maine Environmental Data and Information Management System (EDIMS).* Three U.S. states (Massachusetts, New Hampshire, and Maine) and two Canadian provinces (New Brunswick and Nova Scotia) formed the Gulf of Maine Council on the Marine Environment in early 1990 to facilitate joint use and protection of their shared coastal and marine habitats. One of the Council's main concerns was the interchange of information among its participants, by means of an Environmental Data and Information Management System (EDIMS) (Brown et al., 1994). This System was initially built in the pre-Web years, then lost its funding for a couple of years just as the Web was becoming widespread, and had to regain its legitimacy amidst a very diverse set of organizations.

- *Northwest Environmental Database / Coordinated Information System / StreamNet.* Beginning in 1984, the states and tribes of the Pacific Northwest (Montana, Idaho, Washington, and Oregon) worked with the federal Bonneville Power Administration to build two region-wide rivers information systems, the Northwest Environmental Database (NED), describing river-related fisheries, wildlife, and hydroelectric facilities, and the Columbia River Coordinated Information System (CIS), that tracked anadromous fisheries. Data for both systems were shared via coordinators in each state, using geographic stream identifiers, and encountered a variety of challenges in maintaining region-wide standards and joint usage among agencies with a wide variety of technological maturity. In 1996, both systems were subsumed into an advanced Internet-based data interchange system, known as StreamNet (BPA, 1996).

c. Data collection methods

To compare the information sharing infrastructures to each other, understand their patterns of change, and document new tasks or processes specifically enabled by information sharing, I needed an in-depth view of the choices people had made and were making, along with the context in which these choices were made. I also knew that my chosen topic was a complex, unpredictable one, that could well take unexpected turns as I learned about my cases. Therefore, in keeping with Yin (1984), I opted for loosely structured interviews with the key architects, builders, users, and

---

[2] This list originally included a fourth case, the Gulf of Mexico Program Information Network (Cobb and Seals, 1996); but its infrastructure was the most embryonic of the four, and I ended up dropping it from the list after completing the first three cases.

supporters of the infrastructures. I drew up a three-page interview protocol, in five parts: *(i)* new tasks and processes due to shared information; *(ii)* the size and quality of information sharing; *(iii)* the relationship of information sharing to changed tasks or processes [for *internal validity* (Yin, 1984)]; *(iv)* the infrastructure's technological design and context; and *(v)* its organizational context.

In each case, to select interview participants, I began with written materials sent to me by my contact from my preliminary inquiries. Using these materials (and Web pages where available), I drew up a preliminary list of names from committee rosters, author lists, and meeting attendance lists. I submitted this list to the "informant" in each case, who suggested additions, removals, and corrections; and then I set about contacting the various people on my list, either by electronic mail or by telephone, to schedule in-person, onsite interviews. During initial contacts with these interviewees, several of them in turn suggested additional interviewees. Table 3-1 summarizes the parameters of the three case study efforts.

	Great Lakes Information Network (GLIN)	*Pacific Northwest Environmental Database / CIS / StreamNet*	*Gulf of Maine EDIMS*
Fieldwork	7 days Feb. 17-28, 1995	10 days Apr. 25 - May 5, 1995	6 days May 23 - July 19, 1996
Locations	IL: Chicago MI: Ann Arbor, Detroit, Lansing, Bay City ON: Windsor, Toronto, Etobicoke, Hamilton	OR: Portland, Salem, Gladstone WA: Olympia, Lacey, Seattle MT: Kalispell, Helena ID: Fort Hall, Boise	NH: Durham ME: Augusta, Falmouth, Biddeford NS: Halifax, Dartmouth MA: Boston
Interviewees	17	32	12
Documents collected	7” shelf space[3] 15 lbs.	36” shelf space 78 lbs.	10” shelf space 22 lbs.
Draft review	April 1995	Jan. 1996	Aug. 1996

Table 3-1. Case study parameters

For each case, prior to embarking on fieldwork, I conducted extensive Web searches to familiarize myself with my interviewees' context and tasks as much as possible before meeting them, so as to adapt the interview to their sphere of activity. For each case, I conducted the interviews, 45 minutes to 2 hours each, in a one to two-week automobile tour of the states and provinces involved (for the third case, closer to home, my fieldwork was more drawn out). I used portions of the in-

---

[3] In an increasingly digital age, my need for large, sturdy furniture isn't the whole story: besides armfuls of reports, I collected some information on diskettes, and a lot on the World Wide Web. Still, these numbers are a good measure of how much I had to work with in each case (and they greatly affected the logistics of air travel home from fieldwork).

terview protocol (not all of it applied to each interviewee) to investigate the use and role of shared information in each organization, the apparent factors in its growth or stagnation over time, and the role of technological and organizational contexts and priorities.

These interviews led to additional readings (internal reports, educational brochures, electronic mail archives, meeting minutes, and Web pages); additional telephone and electronic-mail conversations with interviewees; and reviews of databases, geographic datasets, and online data services. I continued doing spot checks of the key Web pages in each case every few weeks, right up until May 1997. As listed in Table 3-1, I sent out draft case summaries for review to many interviewees (those whose information I used) at three points along the way: their responses, mostly by electronic mail, provided accuracy checks and additional insights. Other sources of information included an educational videotape, a few lengthy unstructured conversations, attendance at committee meetings in the Gulf of Maine case, and (especially in the Pacific Northwest case) direct observation of dams, electric power facilities, fish hatcheries, and landscapes (I drove 2,800 miles). All of these measures strengthened the study's *construct validity* (Yin, 1984).

d. Qualitative analysis

Analysis of the study's findings followed methods outlined by Yin (1984) and Eisenhardt (1989) for literal and theoretical replication. I expected the cases to build on each other according to *literal replication*, in which similar findings reinforce the logic derived from earlier ones. I also hoped to draw *theoretical replication* from two dissimilarities between the cases. First, the Pacific Northwest case, in the absence of a data network, would show different outcomes; second, the cases ranged in longevity from about a year to over ten years, possibly suggesting fruitful growth paths over time.

This was my *a priori* strategy; but Yin is rather brief on the difficult step of generating patterns from complex findings. The results from my fieldwork proved complex enough to require additional hypothesis-generating approaches. For this I drew on *grounded theory* (Turner, 1983; Glaser and Strauss, 1967), an inductive approach that generates structural patterns and hypotheses from repeated synthesis of the qualitative findings themselves, within their full organizational context—in contrast with traditional hypothesis-testing methods which seek a more context-neutral, formal model. Indeed, although my initial quest was for a *variance* theory predicting an infrastructure's organizational outcomes based on its technological design, I soon shifted to a *process* theory (Markus and Robey, 1988), to understand the choices made in context as the infrastructures were conceived, designed, built, used, and altered over time. This approach was a better fit to the complexity of the information-sharing phenomenon and the richness of the chosen empirical contexts, which are influenced just as much by individual choices than by a set of factors or conditions. The strength of a process model is that it provides a rich understanding of these choices and behavioral patterns as they unfold over time, which can then be used to predict outcomes in closely similar contexts. The process approach also made good use of the individual views expressed by interviewees (and of the researcher's own partial understanding), instead of discarding them in search of an abstract set of variables and causal relationships.

e. Discussion

This research method presents several potential weaknesses, including its limited generalizability beyond a narrow range of cases, the tension between individual choices and broader trends, and the iterative, never-finished analysis process.

First, some have criticized the qualitative, case-based approach for its difficulty in generalizing beyond a few specific contexts. In response, Yin (1984) argues that cases are not statistical "samples," and that the goal in case study research is to understand behavioral logic ("analytical generalization"), not to enumerate frequencies ("statistical generalization"). Yet this is only a partial response: even behavioral logic patterns that arise from such research are usually only necessary, not sufficient, conditions (Markus and Robey, 1988), so the generalizations provided by this research may still seem rather weak. However, when studying voluntary human decisions and actions, such "soft" predictions are expected and appropriate: social forces, political contexts, or technological resources cannot have a predictable effect on people's choices. Thus, the value of case study findings lies not in their complete generality, but in the behavioral insights they suggest. These insights can be used to build organizational and technological *savoir-faire* for other, similar contexts, rather than universally valid formulas or techniques.

Second, my choice of cases might seem overly narrow. It is indeed skewed towards decentralized, loosely-coupled technological and organizational structures, and may offer few implications for more integrated or centralized structures, more unsystematic, *ad hoc* approaches, or even organizations that are *not* sharing information at all. But my choice was motivated by an interest not just in information sharing *per se*, but in the design and implementation of infrastructures for information sharing between independent systems. To broaden the set of cases, I would have had to stretch the infrastructure concept quite far; I had already loosened my "quality" criteria quite a bit to fit the mid-1990s networking context. Another limitation of my choice of cases is that they all feature a shared natural resource. This choice excluded inter-agency groups organized around metropolitan areas, transportation networks, or states; and local or national cooperative information systems efforts. However, this choice ensured that organizations were linked by a well-defined physical relationship, and it provided a degree of comparability across the cases.

Third, the narrative, context-specific nature of my interviews and findings presented the risk of concentrating on individual decisions, and missing underlying trends. The structuration perspective presented in Chapter 2, with its sensitivity to broad structural shifts, provides some protection against this risk; but the study's conclusions do remain open to debate about the role of individual choices vs. underlying social forces. But here again, this "limitation" is in fact a more accurate representation of human behavior within social systems: both individual choices and broader trends are usually at work, and become more or less visible depending on the level of analysis: individual, meso-, and macro scale (Misa, 1994). Monitoring the cases for some time would help to confirm, correct, and deepen the study's insights on the role of individual choices *vs.* that of broader societal forces.

Fourth, my loosely structured, qualitative approach made it difficult to predict not only the study's findings, but even the nature of those findings: prior to collecting and analyzing the data, I had an interview protocol, but the deeper questions to be answered were themselves undefined. Thus, a single set of interviews provided only limited insight; and subsequent analysis was quite difficult, given that it was not guided by any prior structure or hypothesis. Drawing conclusions from these findings was not a one-time event, but an ongoing process of (re)interpretation—long after formal data collection was complete. This is not surprising given the complexity of the study topic: its conclusions are not only context-specific and not fully certain, but perennially subject to interpretation and debate even within their own context. However, this kind of study is valuable not for its finality, but for its insights into behavioral logic. This logic, though context-specific, is nonetheless likely to hold in other, similar contexts, and can thus be used to build organizational and technological *savoir-faire*. The lessons learned from this study are intended to be cogent contributions to ongoing discussions in the technological, organizational, and policy fields. For this sort of learning, the qualitative, interpretive approach followed here works very well.

3. Prototype development

For the second phase of the research, my objective was to provide a "reality check" on the case-study findings, a technological counterpoint to the case studies' organizational focus, that would help me to extrapolate their findings beyond the current state-of-the-art into the near future within a changing technological context. The next three paragraphs explain the choice of prototype, the *a priori* design choices, and methods for evaluation.

a. Selection of prototype

First, it's hard to prototype an information infrastructure usefully, especially the kind investigated here, in a laboratory setting. However, the US National Spatial Data Infrastructure provided a natural context for this kind of experimentation; and the Federal Geographic Data Committee's seed funding, through the Competitive Cooperative Agreements Program (CCAP), was a good match to the intent of my research, and to the timeline and scope I needed. So, together with a small team of researchers in the Computer Resource Laboratory of MIT's Department of Urban Studies and Planning, I wrote a proposal for CCAP funding in February 1995 (just prior to my first segment of fieldwork). With partial support from these funds, I worked on this project between November 1995 and April 1997, while following up from the first two case studies and conducting the third.

In keeping with the Competitive Cooperative Agreements Program, the project was to create a node on the NSDI's Geospatial Data Clearinghouse—that is, a Web-based service providing a searchable standard interface to geographic metadata. This was part of what I wanted to experiment with; but I also wanted to learn what it took to provide the geographic information itself online. While not a requirement of CCAP funding, this was nonetheless an area of great interest to the NSDI Clearinghouse (Nebert, 1995) and to other members of the team at MIT.

The choice of orthophotos arose from three lines of reasoning: First, orthophotos were gaining prominence in state GIS centers, under the US Geological Survey's digital orthophoto quadrangle (DOQ) program, and they were increasingly seen as an important "Framework" data layer in the National Spatial Data Infrastructure (FGDC, 1995). Yet the usual way in which orthophotos were distributed (very large files on compact discs) restricted their use to a small cadre of advanced GIS specialists—even though the full orthophotos, at full resolution, were rarely what people needed. So it seemed worthwhile to investigate alternate distribution mechanisms for these orthophotos over the Internet (e.g. delivering only the desired piece of an orthophoto, and at a variety of resolution levels) to see whether these mechanisms might make orthophoto data more accessible to a wider audience. Second, given that orthophotos are raster images precisely aligned along geographic coordinate axes, they provided a simple data structure, with simple ties to geography. This would allow the project to explore the design of the data interface, and questions of geographic standards and interoperability, without spending too much time parsing and translating file formats, or transforming the images to fit geographic coordinates.

b. Design choices

Some of the design choices behind the orthophoto prototype, such as standardized, searchable metadata, were fixed by the requirements of the CCAP funding. Other choices went beyond the program's requirements, and were more open-ended. For instance, from the earliest design stages, it was clear that the orthophoto browser had to use the network sparingly, so as to offer a viable alternative to CDs or other conventional distribution mechanisms. This called for distributing small sized orthophoto excerpts, in compressed form. Also, because the information was geographical in nature, it was important to give the browser an intuitive geographic interface, with basic actions such as zoom in/out and pan north/south/etc., but without re-creating an entire mapping package. A third design goal, motivated by an interest in broader questions of geographic interoperability, was to facilitate reading the orthophoto data into client-side GIS or mapping software, to help users integrate local with remote data. A fourth development goal (really an underlying principle) was a staged development path that would define and target intermediate functionality goals, while preparing for more sophisticated capabilities. This led, for instance, to getting a simple interface up and running quickly with fixed image tiles; but the underlying script anticipated a variable tile size and was therefore easy to adapt.

c. Evaluation

Given the nature of the project and my use of it in the context of this thesis, the focus is not on a detailed usage study, but on building and interpreting a "proof-of-concept" to illustrate and instantiate several ideas about geographic information sharing. Nevertheless, one important step was taken to provide valuable usage information: when the orthophoto browser was announced, the Web server's log was restarted and it logged every Web transaction over more than six months. I assessed people's use of the orthophoto prototype in three ways: I perused and analyzed the Web server's transaction log (Sept. 1996-March 1997) through relational database queries and *wwwstat* software; I read and kept the many comments that came in by electronic mail by users of the browser; and I placed telephone calls to a few known users in Boston city government to gather anecdotal evidence about modes of use.

4. Synopsis

As a result of this hybrid study, the structure of the next few chapters may require a brief explanation. Chapters 4, 5, and 6 contain the case-by-case summaries from fieldwork. These write-ups follow a common outline, which facilitates cross-case comparisons and synthesis in Chapter 7. Chapter 8 then describes the orthophoto prototype project, and draws conclusions from both the development process and the final product achieved. Lastly, Chapter 9 draws on the conclusions from both the case studies and the prototype to sketch the implications of effective geographic infrastructures for technology, organizations, and policy.

Chapter 4: The Great Lakes Information Network

1. Overview

The Great Lakes, bordering on eight US states (Minnesota, Wisconsin, Illinois, Indiana, Michigan, Ohio, Pennsylvania, and New York) and the Canadian province of Ontario, comprise the world's second-largest freshwater system (after the polar ice caps), containing about 18% of the world's supply. The Great Lakes basin, some 500,000 km^2 (200,000 sq. mi.) in area, is home to over 37 million people, over 10% of the US population and a quarter of Canada's population (EPA-GLNPO, 1995; IJC, 1995).

Figure 4-1. Great Lakes geography

Despite environmental improvements since the 1960s, the Great Lakes ecosystem remains a matter of concern due to toxic air and water pollutants, nutrient loading from shoreline development, and introductions of non-native ("nuisance," or "exotic") aquatic species (IJC, 1995), all exerting pressure on a largely closed system: for instance, Lake Michigan's retention time is 99 years. These concerns have led to several inter-state and bi-national agreements, and have created "a legacy of co-operation" among the region's state, province, and federal agencies since the 1950s (Ralston, 1995).

The Great Lakes Commission is one outcome of these regional agreements: an interstate compact of the eight Great Lakes states, it was formed in 1955 with a mandate to promote "orderly, integrated, and comprehensive" management of Great Lakes water resources. In mid-1993, the Great Lakes Commission set out to link the region's various stakeholders via networked communications and information interchange. This effort, known as the Great Lakes Information Network (GLIN), quickly enlisted the participation of dozens of U.S. and Canadian Federal agencies, state and provincial agencies, and other public and private organizations with an interest in the ecology and economy of the Great Lakes. The resulting network has gained quite high visibility as a "one-stop shopping" resource for information on Great Lakes-related organizations and activities. As such, GLIN provides an interesting look at the challenges of growing a highly networked, mostly unplanned information infrastructure within a complex institutional context.

The following sections first sketch the history of the GLIN infrastructure, its institutional and technological context, and the choices made in building it. Subsequent sections then evaluate GLIN by the quality criteria outlined in the previous chapter, and by its impacts on organizations and policy. The last section outlines the challenge of drawing on the information resources of participating organizations, and of interacting fruitfully with other, overlapping information-infrastructure efforts.

2. Brief history

> "The little pilot has become a messy beast!"
> - GLIN Board minutes, July 1994

The GLIN concept was launched in mid-1993 by the Communications Director at the Great Lakes Commission (GLC), in partnership with a systems specialist at CICNet (a regional nonprofit Internet service provider established by a consortium of Midwest research universities). Beginning with an extensive survey of potential users and their needs, they secured a $220,000 grant for a two-year pilot study from the Ameritech Foundation (Ameritech is the "baby Bell" telecommunications company that serves five out of the eight Great Lakes states, from Wisconsin to Ohio: its Foundation has a long record of charitable giving in these states.)

The pilot project allowed them to leverage additional grants from the U.S. Environmental Protection Agency (EPA) and the National Telecommunications Infrastructure Administration (NTIA); Ameritech renewed and increased its funding in 1996. In April 1994, an interagency Advisory Board was formed to represent about 20 different organizations in guiding GLIN's growth. As expressed at this first Advisory Board, GLIN's intent was to help organizations all over the Great Lakes put up information servers and to link them to each other: "You will keep the information on the computer in your office—and the Internet, a common tool, will connect us all." (GLIN Board minutes, April 1994).

During the two-year pilot phase, the network attained high visibility, with significant media coverage (WSJ, 1994) and data links from Canada's Environmental "Green Lane," the U.S. EPA, and others. Its use grew more rapidly than anyone expected, and in mid-1995, as the two-year pilot stage drew to a close, the Advisory Board was faced with several challenges for the network's future: building financial and technical self-sufficiency, making effective use of GLIN's growing web of information, encouraging holders of ecological and economic data to put it online, and defining GLIN's scope in relation to other, overlapping data sharing efforts in the region. The next several sections explain how they arrived at these challenges, and suggest some possible solutions.

3. Context of the infrastructure

a. Institutional context

In devising GLIN, the Great Lakes Commission was building on a long tradition of cooperation among the states and with federal agencies and others in the region. Past agreements (rules, treaties, and the like) on the Great Lakes environment include *(i)* the Boundary Waters Treaty of 1909 (water use and allocation between the US and Canada, overseen by the International Joint Commission); *(ii)* the Great Lakes Water Quality Initiatives of 1972 (phosphorus runoff), 1978 (toxic pollutants), and 1987 (airborne pollutants); and *(iii)* the Great Lakes Binational Toxics Strategy of 1997 (joint US/Canadian pollution targets). The two federal agencies responsible for protection of the Great Lakes are the US Environmental Protection Agency (EPA) and its counterpart Environment Canada. Also at the federal level, the US Army Corps of Engineers, Detroit District, is responsible for maintaining shorelines and shipping channels in the Great Lakes. The governments of the eight Great Lakes states and the province of Ontario form another important part of GLIN's institutional context. Finally, several regional and professional associations (Great Lakes Protection Fund, Council of Great Lakes Governors, and others) have a long-standing interest in the Great Lakes, along with non-profit groups (National Wildlife Federation, Nature Conservancy), and academic and research facilities (the Great Lakes Environmental Research Laboratory of the National Oceanographic and Atmospheric Administration (NOAA)). These few examples suggest the organizational complexity within which GLIN was formed.

b. Technological context

GLIN came into being amidst a significant amount of ongoing information technology effort in the Great Lakes region. Several groups had begun to build sizable information resources on the Great Lakes; and a few had embarked on interagency information sharing infrastructures of their own. The following are salient examples:

First, the Environmental Protection Agency (EPA) had lengthy experience with environmental databases, dating back to the 1960s, with STORET, "the world's oldest and largest environmental data system," running on an IBM mainframe (EPA, 1997). An important EPA emphasis in the Great Lakes was a data warehouse named Envirofacts, containing information on air and water quality, regulated facilities, permits, and Superfund sites. In 1994, EPA staff began building several Internet gateways to Envirofacts (EPA, 1997b; Adhikari, 1997), and integrating it with custom mapping tools ("SiteInfo") on the Web (EPA, 1997c). Furthermore, EPA's regional Water Division had a large set of digital geographic data on the Great Lakes, with plans to make it publicly available via the Internet (EPA, 1996).

The EPA also began a multi-year contract in 1993 with the Consortium for International Earth Science Information Network (CIESIN—pronounced "season"), a non-profit research organization largely funded by the global climate change research program of the National Aeronautic and Space Administration (NASA). Under this contract, CIESIN was cataloging the environmental data holdings of the EPA and other agencies in the Great Lakes region, and building a "Great Lakes Regional Environmental Information System" that would allow users to tap into the EPA's information services via the Internet (i.e. via the Web as well as a special-purpose "CIESIN Gateway" client). (CIESIN, 1996b).

The U.S. Army Corps of Engineers, Detroit District, had a large GIS facility (as of March, 1995, about 7 Gb of vector data and 320 Gb of raster data, including a complete set of digitized shoreline airphotos for the Great Lakes) which they used for shipping lane and shoreline operations and for analyzing the impacts of Great Lakes water levels. The Corps had begun putting real-time hydrological information online (at *http://sparky.nce.usace.army.mil)*, and was working with both GLIN and CIESIN staff to catalog its data holdings and provide networked access to them. In mid-1995 the Corps' headquarters put a server online *(http://corps_geo1.usace.army.mil)* providing Web and WAIS access to metadata on their geographic data holdings.

The Corps also worked closely with the State of Michigan's Department of Natural Resources, which had about 30 Gb of GIS data files, and distributed them via a statewide consortium known as IMAGIN ("Improving Michigan's Access to Geographic Information Networks"), as well as through sales and cooperative agreements. IMAGIN was a carefully institutionalized consortium of public and non-profit organizations in Michigan, funded by the W. K. Kellogg foundation, with formal agreements for exchanging and using geographic information held in a central data archive at the Library of Michigan (Beaulac et al., 1994).

Lastly, Environment Canada and the Canada Center for Inland Waters had begun building the Great Lakes Information Management Resource (GLIMR), a set of environmental data accessible through the World Wide Web and indexed through a multi-server keyword system. Officially launched in March 1995, GLIMR (at *http://www.cciw.ca/glimr)* was the first of several such systems within Environment Canada's "Green Lane" networking initiative.

As a result, GLIN's efforts in information sharing had many precedents among the region's government agencies. This meant that plenty of expertise was available to them; but on the other hand they didn't have the luxury of starting from a clean slate: several competing styles and affiliations for information sharing already existed. (The three information sharing initiatives above (CIESIN, IMAGIN, and GLIMR) formed an unusually important part of the GLIN story: Section 7 below explains how they made GLIN's context both fruitful and chaotic, and suggests the benefits that GLIN might have reaped from a more intentional collaboration with these information sharing efforts.) One final, important element of GLIN's technological context: many state agencies in the region had little or no experience with Internet access, and were eager to get online through GLIN.

4. Infrastructure choices

Partly in response to its institutional and technological context, GLIN embodied several choices of organizational structure and technology.

a. Institutional arrangements

GLIN embraced its complex institutional context from the day it was launched. According to an attendee list on the GLIN gopher server (*gopher.great-lakes.net*), the Great Lakes Commission invited over 100 people to a "GLIN kickoff meeting" in June 1993, representing 60 different organizations. The Advisory Board, formed a year later, was quite broadly defined and represents GLIN's institutional context quite well (Table 4-2).

State / Province	Indiana Department of Natural Resources
	Illinois State Library
	Michigan Department of Natural Resources
	Minnesota Department of Administration
	New York State Department of Environmental Conservation
	Ohio Department of Natural Resources
	Ohio Environmental Protection Agency
	Pennsylvania Department of Environmental Protection
	Wisconsin Governor's Office
	Council of Great Lakes Governors
	Great Lakes Commission
	Ontario Ministry of Environment & Energy
Federal	International Joint Commission
	Great Lakes Fishery Commission
	Environment Canada (incl. Canada Center for Inland Waters)
	National Oceanic and Atmospheric Administration (NOAA), Great Lakes Environmental Research Laboratory
	U.S. Army Corps of Engineers, Detroit District
	U.S. Environmental Protection Agency (EPA), Great Lakes National Program Office (GLNPO)
	U.S. Federal Reserve Bank of Chicago
	U.S. Geological Survey
	U.S. Congress
Tribal	Chippewa/Ottawa Fishery Management Authority
	Great Lakes Regional American Indian Network
Foundations	Ameritech Foundation
	W.K. Kellogg Foundation
Private / Commercial	Council of Great Lakes Industries
	Merit/Michnet
University/ Research	Great Lakes Sea Grant Network
	University of Guelph
	Consortium for International Earth Science Information Network (CIESIN)

Table 4-2. Organizations on the GLIN Advisory Board (May 1997)

GLIN's primary partners include the Great Lakes National Program Office of the U. S. Environmental Protection Agency (EPA), the US Army Corps of Engineers (Detroit District), and Environment Canada.

b. Technological design

In mid-1993, GLIN began as a gopher server, the Internet technology *du jour* at the time. In mid 1994, as the World Wide Web gained prominence on the Internet, GLIN began creating a few Web pages, but didn't make the Web its primary mode of information sharing until late 1994. (In early 1995 the gopher pages still provided more complete information, and any comprehensive search had to include both information spaces.) The Great Lakes Commission intended this Web/Gopher server as a single access point to information on the Great Lakes, but not as an information "hub" in the usual sense: although the Great Lakes Commission's server hosted the pages of several GLIN partners, decentralization was to increase as partners put their own servers online and move their files onto their own machines. When GLIN partner agencies did have a server online, the GLIN front page at *www.great-lakes.net* maintained hypertext links to it.

Much of GLIN's early efforts were focused on helping and training less-experienced agencies to set up their own servers. To get basic communications started quickly, however, many people (about 400, well in excess of the 30 initially expected) were given accounts on the Great Lakes Commission's server, which they reached via dial-up. (This resulted in a greater-than-expected training and technical support burden on GLIN staff: in early 1995, GLIN's technical director expressed concern about having "soft-pedaled" the actual cost by providing as much technical support as they did to these agencies, instead of helping them sooner to rely on local Internet services and support.)

Less visible, but equally important forms of information sharing on GLIN included file transfer protocol (*ftp*) servers and several electronic mailing lists: for instance, one public list, *glin-announce*, kept a broad community abreast of regional events; another, closed, list, *air-toxics*, allowed a few dozen specialists to collaborate on air-pollution monitoring, as part of the RAPIDS project described below.

5. Information sharing characteristics

The information infrastructure that resulted from these choices was quite informally structured, with mostly fixed documents and images, with a few more detailed data sets emerging in early

1995. The following paragraphs relate the characteristics of the infrastructure, including its size and quality.

a. Forms of information sharing

As of early 1995, GLIN's information sharing was of four kinds. Most prominent were summaries (reports, fact sheets, brochures, etc.) that presented quick facts on the Great Lakes and its stakeholder organizations. According to GLIN's director at the Great Lakes Commission, the audience for this information was the "intelligent layperson," including staff in other public agencies, educators, and the general public. Electronic mailing lists and sharing of data and software through *ftp* were another form of information sharing on GLIN. The result was akin to an atlas or reference library, with a lot of facts within easy reach, and coming from many reliable sources.

Yet many respondents to the pre-GLIN user survey in 1992 had expressed the need for "technical data" and "research findings" on a frequent basis; and in early 1995, a few detailed, dynamic data holdings began to come online: the Great Lakes Environmental Research Laboratory had daily satellite images from the "CoastWatch" program of the National Oceanographic and Atmospheric Administration (NOAA); and the US Army Corps of Engineers had begun putting reports and forecasts of Great Lakes water levels online. (Interestingly, not everyone expected their detailed data to be of much interest to anyone else. A member of GLIN's Advisory Board recalled that federal agencies invited to a GLIN meeting in September 1993 conceded that they could make data available to GLIN with some effort, given that it "belonged to the scientists"—but that "no one would likely be interested.") The Great Lakes Commission had always had an interest in providing detailed geographic data from the start; but as of early 1995, GLIN's forays into distributing GIS data were still limited to one person's experiments with Arcview software. (However, Commission staff continued their experimentation; in mid-1996, as off-the-shelf toolkits became available for integrating GIS into Web services, they evaluated MapInfo Corp.'s ProServer and ESRI's ArcView Internet Map Server. As a result, in May 1997 the Commission was invited by the EPA's Great Lakes office to submit a proposal for a four-year project to tie GLIN to networked GIS services.)

b. Size of the infrastructure

There are many ways to measure GLIN. Some participants saw size or usage data as relatively unimportant; but others had carefully assessed the number of Web pages on their server, the number of users, the size of their data holdings, etc. In early 1995, the *www.great-lakes.net* server had 250 user accounts, 500 email list participants, and some 1,400 files online. In February 1995, the gopher server (*gopher.great-lakes.net*) reached a peak of 12,600 file retrievals a month, before declining gradually as Web access became more widespread. In April 1997, the Webserver at *www.great-lakes.net* received 67,000 "hits" (requests), transferring a total of 442 Mbytes. These are clearly only rough measures: not all hits represent Great Lakes users; there are always more transactions than actual pages retrieved; and hits tend to lump together large and small files. GLIN's director at the Great Lakes Commission typically paid more attention to another measure of growth, the rate at which state agencies in the region were putting information servers online. This showed much slower growth, so she concentrated her efforts there.

c. Quality of the shared information

As outlined in the previous chapter, the quality of an information infrastructure depends on how well the shared information matches the needs of its users, in particular its accuracy, timeliness, and encapsulation.

i. Precision and accuracy

This was frequent concern about networked information among GLIN participants: "will the search for information be worth my time?" asked a systems analyst at the International Joint Commission. (For similar reasons, a GIS manager at the Environmental Protection Agency had stopped providing his staff with tools to access the Internet unless they had a specific purpose in mind.) To ensure the quality of GLIN's information across a network of data sources, GLIN staff concentrated on bringing providers of high quality information online: "we will not point to or work with just anyone," said GLIN's director. GLIN staff spent a lot of time reviewing pages on *www.great-lakes.net* for quality control (bad links, etc.); at the bottom of each page they would leave a "time-stamp" indicating who last updated the page and when.

ii. Concurrency and timeliness

Much of the information shared via GLIN consisted of agency fact sheets, laws and policies, and other static data: concurrency and timeliness were not a prime concern for these. One source of real-time information was the Great Lakes water levels forecasts by the Army Corps of Engineers. Other changing information on GLIN included event calendars: these were checked and updated frequently by dedicated staff at the Great Lakes Commission. Interestingly, several early descriptions of GLIN emphasized the need for electronic networks in sharing rapidly-changing information, arguing that "the shelf-life of Great Lakes information has shortened as the volume, diversity and the need for quick access to it have increased dramatically." When Ameritech made its initial grant to GLIN, the announcement stated that GLIN would "provide quick access to today's information, not yesterday's or last week's." Yet as of early 1995, concurrency and timeliness of information weren't yet a concern for most GLIN participants, given that most GLIN documents were semi-permanent fact-sheets or records of some kind.

iii. Usability and encapsulation

At one level, sharing information on the World Wide Web makes encapsulation trivial: Web browsers can interpret the Web's Hypertext Markup Language (HTML) regardless of computer or operating system. But going beyond static HTML text and graphics requires additional thought as to how easily the audience can correctly use the data or service. For the most part, most information shared on GLIN was simple text and graphics. Encapsulating more complex data came a good bit later, in 1996-97, as the Great Lakes Commission staff explored mechanisms to provide interactive GIS services to Great Lakes users via the Web, using networked add-ons to MapInfo and ArcView GIS software.

d. Quality of the information infrastructure

The above quality measures are from the perspective of the data recipient. Several additional "quality" aspects of information sharing pertain not to the data, but to the infrastructure that provides it.

i. Reciprocity: what it takes to publish data

Anyone with a direct Internet connection can publish things via Gopher or the Web; in that sense, GLIN data sharing was reciprocal. However, most files on GLIN came from a few heavy-duty information providers; most GLIN participants were merely data and document recipients. GLIN's technical director, along with GLIN's representative from EPA's Great Lakes National Program Office, expressed disappointment at how slowly GLIN partners had put their own information online. They had hoped that easy access to data across the region would prompt many to quickly publish their own data. Other GLIN participants took a different view: one member of GLIN's Advisory Board was encouraged by those who had contributed, and expected many more to follow suit.

ii. Flexibility: how people join the network

Technologically, the gopher and Web topologies made growing GLIN very easy: Great Lakes Commission staff invited the participation of known providers of high-quality information, helped them to put a server online, and put links to the new site on the main GLIN pages. Training was the primary burden, mostly felt by GLIN's technical director, who provided training and technical assistance to an increasing number of still-novice users. The Great Lakes Commission had limited training resources, which it focused on promoting free and open data communication among its eight constituent states and related federal agencies. This policy tended to exclude small non-government groups and local communities, but it was in keeping with the Great Lakes Commission's role as an inter-state compact.

iii. Scalability and sustainability

From its outset, GLIN was always described as a network of data providers, rather than a data center. Accordingly, the small staff at the Great Lakes Commission and CICNet encouraged GLIN partners to become "self-sufficient" by teaching them to take full responsibility for providing and maintaining their own information on their own servers. The goal was a self-supporting infrastructure that would not depend on any one "hub" server. However, according to GLIN's technical director, its early focus on outreach tended to under-represent the actual cost of Internet connectivity (the time to train, the required networking investments), and produced a large group of

Internet novices with dial-up connections to *www.great-lakes.net*, instead of their own Internet node, who had made minimal data contributions to GLIN, and who continued to require a lot of technical support. In the interest of sustainability, by mid-1996 GLIN staff required (and facilitated) people's use of local commercial Internet services, and they encouraged states to officially "endorse" GLIN through a formal resolution and budget items for networked information sharing. This latter thrust in particular was a departure from the informal, good-faith agreements that GLIN was raised on. Greater formality was needed in part because many state-level contacts had limited decision-making influence in their agencies, and had to rely on more formal processes to persuade their agencies to invest in contributing to GLIN.

iv. Non-intrusiveness and transparency: how much change is required

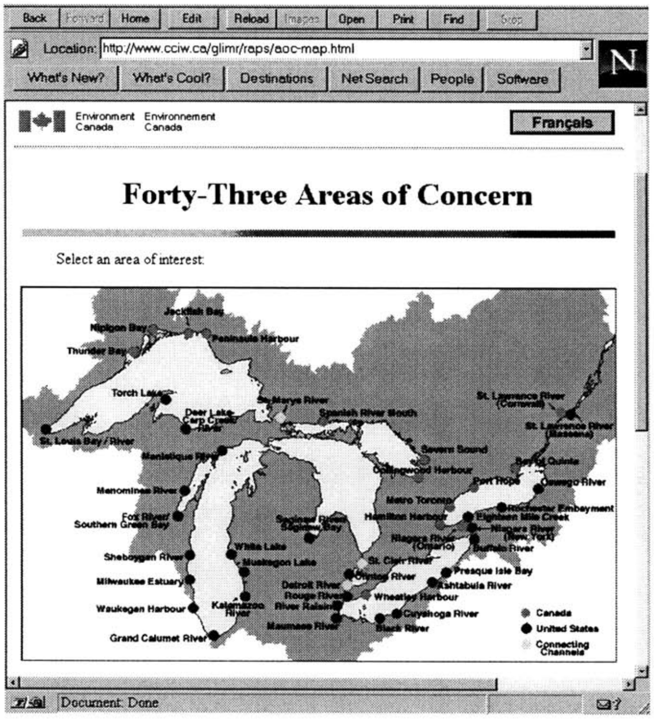
GLIN imposed few restrictions on what people could put online: "I don't care what they publish," said GLIN's director; and the gopher and Web servers did indeed provide access to many different kinds of information. But as GLIN's technical director pointed out, putting data online does imply significant restrictions on the data: putting a FileMaker database on the Web requires data conversion expertise that was beyond the reach of most GLIN participants. (Interchanging detailed scientific and/or geographic data would have raised a number of other standards and compatibility questions; but these hadn't yet surfaced in early 1995.)

6. Infrastructure impacts

Finally, perhaps the most important measure of an information sharing infrastructure is its impact on people and organizations. GLIN's potential was always clear. To one EPA data specialist, it offered a chance to see other people's information for free and in a "one-stop shopping" mode. For a researcher at the Great Lakes Environmental Research Laboratory, it afforded opportunities for publicizing current laboratory experiments and finding other researchers to work with. To the EPA and many others, it provided a vital public relations function. GLIN's success stories often revolved around its high-visibility users (high-ranking officials), its inclusion of many diverse participants, or its rapid rise in usage, beyond its founders' expectations.

a. Interdependent, "ecosystem" thinking

One important impact of GLIN on people and organizations was the interdisciplinary "ecosystem perspective" it presented on the Great Lakes. GLIN staff sought out information providers from a wide variety of sectors: pollution prevention (EPA), industrial development (the Council of Great Lakes Industries), business and economics (Federal Reserve Bank of Chicago), and nature conservation (National Wildlife Federation), to name a few. Putting all these kinds of information near each other, linked by hypertext, may have highlighted the complexity, diversity, and interdependency of all these parts of the Great Lakes ecosystem. (However, as one researcher at Environment Canada observed, this "ecosystem thinking" effect depended on the user: some might simply "raid" GLIN to get what they wanted regardless of GLIN's educational intent.) A good illustration of how GLIN could communicate Great Lakes issues in a holistic way is the interactive map of "Areas of Concern" on the Web (Figure 4-3), a page jointly maintained by Environment Canada and the US EPA. On this map, color-coded dots representing harbors or shipping channels identified as needing remedial action by the 1987 Great Lakes Water Quality Agreement. Each of these dots is linked to a Web page on either EPA or Environment Canada servers, depending on which side of the border it's on. This simple artifact communicated powerfully and creatively the interdependence of the two country's actions, and emphasized the cooperation between counterpart environmental agencies (insofar as making joint maps indicates cooperation).



Forty-Three Areas of Concern

Select an area of interest:

Figure 4-3. Areas of Concern joint Web map

This begs the question: has GLIN actually changed any tasks or processes of environmental management? Tangible evidence is still too scarce to tell. GLIN may still be too new a facility for people to depend on it for their work, and to change their work processes accordingly. Another view is that GLIN's loosely organized growth to date has made it difficult to find needed information across its many data sources. GLIN staff point to one clear impact, the Rapid Air Pollutant Inventory Development System (RAPIDS).

b. RAPIDS: a clear impact?

RAPIDS was a complex, science-intensive piece of software developed by air-quality specialists and software engineers around the Great Lakes region. Contributors communicated with each other via the *air-toxics* mailing list housed on *www.great-lakes.net.* They used this list to discuss various aspects of the analysis, and to provide feedback on successive versions of the product, which they shared via the *ftp* server at *ftp.great-lakes.net* (electronic exchange of software enabled new versions to be distributed to the whole group in minutes). Another link with GLIN was that RAPIDS was coordinated by GLIN's director at the Great Lakes Commission. Many involved felt that the RAPIDS product was of exceptional quality and that it was produced exceptionally quickly; and they attributed both of these aspects to the networked collaboration that went into developing it, and it became known as a clear impact of GLIN. However, RAPIDS was a separate effort from GLIN, with its own budget, personnel, and goals; its only links to GLIN were the machines it used as mailhub and software locker, and its coordination by GLIN's director. So, while RAPIDS does show the impacts of collaboration over the Internet, its role as a GLIN success story is not nearly as clear. It does, however, suggest the kind of joint projects that *could* occur through the use of GLIN's diverse information and expertise.

7. Challenges and lessons

By any standard, GLIN was an ambitious project, and it surfaced several dilemmas related to supporting its growth, populating the servers with detailed environmental data, and defining the network's purpose amidst a complex network-of-networks.

a. Supporting GLIN's growth

As the GLIN community grew, so did the burden on its staff. In particular, the growth of user accounts on *great-lakes.net* had a drastic impact on GLIN's technical director. As the lone technical contact for all current and potential users, he expressed concern that GLIN staff were becoming, in his words, "a bottleneck to the network." For instance, supporting such a large user community had kept him from launching longer-term GLIN projects such as indexes, interfaces, improved search strategies, or anonymous electronic discussions. He felt that GLIN staff should have been more strict about (merely) training and equipping GLIN partners, and about pushing them to become real Internet nodes with local support, instead of providing so many with Web and Gopher files and accounts on *great-lakes.net.*

b. Populating GLIN with "real" data

One of GLIN's primary goals was to bring together data from many different sources in the region. However, holders of detailed information were slow to publish it on the Internet. Some were wary of Internet security breaches; some saw little outside demand for their data; some were concerned about losing control over their data; still others were concerned about the staff burden of putting data online and providing support to users.

i. Security concerns

The Internet was an unknown territory to many data managers in the region. As of early 1995, GLIN's technical director in particular was surprised at how many technical staff, including many Information Systems managers, were unfamiliar with the Internet. Citing people's fear of being "contaminated" by the Internet, he recalled one state employee who, after using his laptop to dial into an Internet service, was forbidden to connect it to his department's mainframe. Some data providers in the region saw Internet connectivity as bringing many benefits, but still had concerns about security, and were cautious about going online in the absence of a statewide Internet implementation plan—otherwise, said one state GIS manager, "we'd be alone in figuring it out." Members of GLIN's own Advisory Board voiced concerns (either their own or those of others in their agencies) about data being altered once it appeared on a Web or Gopher server. The then-chairman of the GLIN Board suggested that decisive leadership from the top (e.g. state governors) would be crucial to building up Internet connectivity among state governments.

ii. No perceived need

Another impediment to putting data online was the perception that no-one was interested. This was the prevailing view at GLIN's meeting with Federal agencies in September 1993: several offered to publish press releases or meeting announcements, but few saw any reason to compile actual scientific data and put them online. The first chair of GLIN's Advisory Board saw it as his mission to persuade Federal agencies in the region to adopt a policy of openness to collaborators and to the public. He pointed out that free and open sharing was a high priority for both the U.S. Democratic Administration and the Republican party; and besides, he argued, freely shared information "provides the fuel for people's intellect," and withholding it would prevent people from devising creative solutions to problems, resulting in significant costs to society.

iii. Risk of misinterpretation or misuse

Third, several would-be providers of detailed data were concerned about giving up control over their data. A few GLIN partners in Environment Canada found that the Gopher and Web models alleviated this concern by letting providers maintain their own data. Many holders of data were concerned about its correct interpretation and proper use by a wide variety of unknown users. As one researcher at the Great Lakes Environmental Research Laboratory put it, "when specialists who are known to you are using your data, they know *(i)* what to do with it; and *(ii)* how to call you up and ask you about it." This "sheltered" sharing wasn't possible on a network like GLIN; so several GLIN participants took to adding a timestamp and point-of-contact signature to every piece of data put online. Finally, some holders of information preferred that their information not fall into the wrong hands, and so they restricted access to not-for-profit groups, or shared information freely but only among institutional partners.

iv. Staff burdens

Fourth, some providers (and would-be providers) were concerned about the staff burden implied by putting large volumes of data online. One EPA staff member described himself as a "sea anchor" for that agency's data distribution efforts, the one to insist on carefully documenting and efficiently managing geographic data before releasing them to outside users. He felt that without this precaution, his small GIS staff could be swamped with user questions. "We will release no data before it's time!" quipped a hydrologist at the U.S. Army Corps of Engineers in Detroit, who managed the large GIS shop there. His philosophy: *(i)* automate quickly; *(ii)* provide complete metadata; *(iii)* coalesce the data into an efficient distribution medium. Others, however, were less concerned: a data manager at EPA wanted his agency to release documented GIS datasets via *ftp* as soon as possible; GLIN's director reported a relatively low volume of user questions or requests; and an Environment Canada manager was confident that even a large volume of user questions could be handled with the right communications technology.

c. Adjusting to a complex network-of-networks

GLIN was initially conceived as a single autonomous network, but its content and purpose overlapped with several other ongoing efforts to share environmental data in the Great Lakes region. One of its more difficult challenges was how to define its role in light of these other, significant information-sharing projects, some of which had quite different styles and views of information sharing: Environment Canada's Great Lakes Information Management Resource (GLIMR), EPA and CIESIN's Great Lakes Regional Environmental Information System (GLREIS), and the State of Michigan's IMAGIN ("Improving Michigan's Access to Geographic Information Networks") consortium.

i. Distinct approaches to data sharing

Three different information-sharing infrastructures were being built within the Great Lakes region alongside GLIN in 1993-1995.

First, Environment Canada's Great Lakes Information Management Resource (GLIMR) was essentially GLIN's Canadian counterpart, in that it sought to interchange information on the Canadian side of Great Lakes with anyone on the Internet at no cost. It targeted Environment Canada's offices as sources of information, but it was also creating partnerships with other agencies to link to their information. Rather than overlap with GLIN, it complemented GLIN's focus on the eight Great Lakes states by covering the remainder of the basin; furthermore, GLIMR's chief technical architect was developing multi-server search schemes and shared his technical expertise freely with GLIN staff.

Second, the IMAGIN consortium, led by Michigan's Department of Natural Resources, had a carefully institutionalized policy for sharing geographic data produced by that Department, and for ensuring its appropriate use by (only) Michigan government agencies (local, county, state, federal), public universities, and some non-profit groups. Yearly dues ranged up to $450; members would agree not to "misuse" the information (i.e. use it in a way that would misrepresent its accuracy, or use it for personal or private gain). For those outside the consortium, there was a detailed price list: $120-$900 for geographic datasets, depending on size, and $50 per county for tabular data. A primary concern was to protect the ownership of information and to ensure returns on investment for Michigan's Department of Natural Resources, the creator of the data. Yet IMAGIN's manager preferred not to charge money, but to make agreements for enhancing the information in return for its use. For example, he experimented with free distribution to some engineering firms engaged in public-sector contract work. Nonetheless, some in the region saw these rules as restrictive or expensive: for instance, a GIS specialist suggested that obtaining IMAGIN data for use on GLIN would be difficult.

Third, CIESIN's Great Lakes Regional Environmental Information System (GLREIS) was a Web-based system that provided access to metadata, data, and data utilities of relevance to the Great Lakes region." Part of building this system involved cataloging the region's data holdings (primarily EPA's, but those of other data providers as well, such as the Army Corps of Engineers Detroit District and Michigan's Department of Natural Resources). Within this project, CIESIN also provided networked access to EPA's Oracle databases (ENVIROFACTS and others), modeling and decision support capabilities on the Web, interactive map servers using ESRI's Arc/Info software, and demographic data viewer for Census data. As part of their cataloging effort, CIESIN created simple Web pages describing several agencies in the region (e.g. Michigan DNR and the International Joint Commission). Initially, some in the region mistook CIESIN's work as a shortcut to getting online; but CIESIN's contract with EPA left little room for technology transfer to the region's agencies, and the contract's short time horizon meant that CIESIN couldn't afford to wait for agencies to put their own servers online before cataloging and describing their information. So most of the datasets and guides delivered by CIESIN were on a single server, (*epawww.ciesin.org*), and it had little impact on the network capabilities of agencies in the region. This led some GLIN participants to think CIESIN was promising easy Internet solutions instead of helping clients to build their own. However, CIESIN's agenda was to provide networked access to EPA data and EPA-related data on the Great Lakes, independently of the agencies themselves. Many GLIN participants had long recognized GLIN's need for additional "hard data" of this sort; however, some of them saw CIESIN not as a data-intensive complement to GLIN, but as an overlapping, possibly competing regional information infrastructure, and thus a matter of some concern. According to minutes from the GLIN Board's April 1994 meeting, GLIN's representative from EPA's Great Lakes office "appealed to the Advisory Board to help EPA manage this effort, to make sure that CIESIN's effort is something worthwhile and not redundant with GLIN." In an effort to ensure that the two efforts would be coordinated and mutually beneficial, not long after that meeting the GLIN Board circulated a proposed memorandum of understanding between GLIN and CIESIN, that called for a "common public access mechanism (...) to GLIN and the Great Lakes REIS under the GLIN banner." No such mechanism was ever put in place, and the two programs existed alongside each other, with a few hypertext links to each other but a fair amount of overlap, for instance between GLIN's agency pages and CIESIN's organization guides, or between GLIN's work with geographic data services vs. those that CIESIN had built over a year earlier.

ii. Reconciling the different approaches

What's interesting about these alternative views of information infrastructures is that, first, each one had a "vision," or "creed," that defined its approach to information sharing. GLIN's (and GLIMR's) vision was training agencies to provide information online so that all might reap the benefits of the resulting large set of information. IMAGIN's priority was to formalize the sharing and use of geographic information in Michigan, so as to protect its ownership and appropriate use. CIESIN's vision was to provide Great Lakes metadata, information, and data services to the Internet, for the benefit of EPA, the region, and ultimately the globe. Second, each approach to information infrastructures had one or more "prophets" or "evangelists" that campaigned far and wide to proclaim the creed and enlist followers. GLIN's director and the first chair of the GLIN Board took every opportunity to proclaim their vision for shared information; the manager of Michigan's Resource Information System had a major role in crafting IMAGIN's carefully specified data-sharing agreements; and CIESIN had an environmental programs coordinator who established relationships with key data providers. Third, each of the three had "miracle workers" who provided tangible evidence for the truth of the creed by working hard to turn out the promised outcomes. GLIN's technical director worked hard to help partners get their own sites online, and Great Lakes Commission staff maintained hundreds of Web pages on the main GLIN server. IMAGIN was supported by a large GIS center at the Michigan Department of Natural Resources, where over 20 GIS staff worked two shifts a day. CIESIN had a large research and development laboratory devoted to experimenting with networked data services and data management. Despite these patterns, however, these differing approaches are *not* incompatible "religions," but quite possibly complementary approaches. The more useful questions were not what a "real" partner should be, or "true" sharing; but how to harness available resources to perform along a chosen set of objectives. In particular, GLIN and CIESIN represented two different, but important aspects of information sharing infrastructures, the organizational relationships and autonomy, and the data services and search strategies that would help it to harness the region's collections of environmental information more effectively. By not accommodating CIESIN's different style of information sharing, GLIN may have missed an opportunity to build the dynamic set of "short shelf-life" data it had spoken about in its initial funding proposal. (And conversely, by leaving aside the organizational aspects of linking the region's information, CIESIN may have missed opportunities to build a more distributed, organic network of autonomous data sources.)

Members of GLIN's Advisory Board had suggested making GLIN's focus more specific, cautioning that "we can't be all things to all people." Others were more reluctant to narrow GLIN's scope, suggesting that "GLIN may not define Great Lakes information, but merely deliver it." There was some debate as to whether GLIN's focus was congruent with the GLC's charter, or whether it "may leave the Great Lakes Commission and become more region-wide." (This was resolved for the moment in mid-1995 with a formal Great Lakes Commission Policy Statement endorsing GLIN.) Whatever form GLIN takes in the future, adapting to an increasingly networked environment may require a careful evaluation of GLIN's mission as part of a larger system of Great Lakes infrastructures: what it does best and how that complements, builds on, or reinforces these other data sharing efforts.

8. Conclusions

As a pioneer in networked information sharing among governmental agencies, GLIN made impressive strides in a short time. It was committed to modern networking technology and to linking loosely coupled organizations. Its regional focus on the Great Lakes watershed provided a powerful shared concern among a wide array of participants spanning all levels of government in two separate countries. Yet it had more trouble than expected drawing on the information resources of data providers in the region, and it unexpectedly faced the challenge of other, competing views of information sharing. Only after several years of existence did it begin tackling the question of networked geographic data. As such, GLIN will continue for some time to be a fruitful example of the challenges and tradeoffs of building infrastructures for sharing geographic information.

Chapter 5: The Gulf of Maine Environmental Data and Information Management System

1. Overview

The Gulf of Maine is a semi-enclosed sea in the northwest Atlantic, whose rich coastal and marine ecology sharply defines the culture and economy of three U.S. states and two Canadian provinces. For several years, these five governments have worked on joint management of the Gulf of Maine's natural resources through a bilateral Council on the Marine Environment. One of the Council's stated goals is communication and information inter-change: at its outset in 1990, it created a Data and Information Management Committee (DIMC) to oversee communications and information interchange. One of this Committee's main accomplishments was a clearinghouse for data held by the region's various public agencies and research institutions: the Gulf of Maine Environmental Data and Information Management System (EDIMS). The EDIMS system was intended as a distributed information resource, which would draw on autonomous data sites throughout the region via a common network. This was an ambitious design for the time — especially since a basic networking infra-structure to support it did not become widespread until the recent growth of the Internet's World Wide Web. Although EDIMS has thus far had only a minor impact on the Council's activities, it has served as a rallying point for the Council's data-related activities, and has begun to see some use in supporting communication and collaboration across the region.

Several barriers have kept EDIMS from playing a more significant role: the Council's weak institutional position and insecure funding, the difficulties of coordinating participants with vastly different levels of technical experience, the absence of an agreed-upon networking infrastructure, the lack of a clear shared objective, and little or no sense of inter-dependency among its participants. This complex brew of factors seems thus far to have kept EDIMS from growing beyond an initial centralized proof-of-concept into a distributed and widely-used tool for regional planning and policy-making. In the second half of its ten-year Gulf of Maine Action Plan, the Council on the Marine Environment is beginning to set forth a few clearly shared, information-intensive goals. Time will tell whether this shift can provide the momentum, resources, and direction EDIMS needs to fulfill its promise as an advanced, multi-organizational information infrastructure.

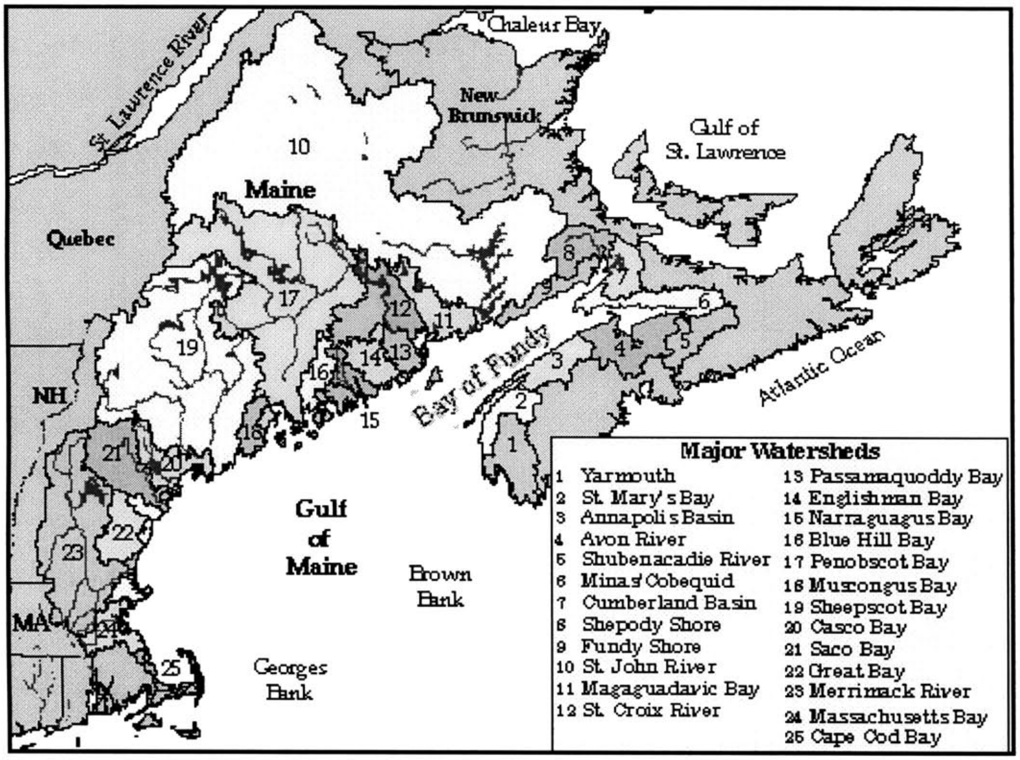


Figure 5-1. Gulf of Maine geography

2. Brief history

In late 1989, the states of Massachusetts, New Hampshire, and Maine, and the provinces of New Brunswick and Nova Scotia, formed the Gulf of Maine Council on the Marine Environment, with a mandate to maintain and enhance the quality of the Gulf of Maine's coastal and marine resources. One of the Council's focus areas was the interchange of information among its participants: in its ten-year Action Plan (1991-2000), it called for "methods to ensure that Gulf environmental databases are compatible," and for the development of a "common regional protocol allowing for the transfer (...) and periodic updating of data and information." To guide these activities, it assembled a Data and Information Management Committee (DIMC) to oversee an Environmental Data and Information Management System (EDIMS) for the Gulf of Maine. This Committee was initially chaired by a member of the Massachusetts Department of Coastal Zone

Management, who secured a grant from the U.S. National Oceanographic and Atmospheric Administration (NOAA) for the Committee to compile a directory of coastal and marine data for the region through a survey of state, province, and federal agencies. Another Committee member, representing the U.S. Environmental Protection Agency, packaged this directory as a Clipper database application; it was then placed on a Unix server at the University of New Hampshire's Ocean Process Analysis Lab, under the supervision of the Lab's Director (who took over as co-chair of the Committee when the founding chairman had to resign by request of his home agency in Massachusetts). At the University of New Hampshire, EDIMS was coordinated by a Research Associate in the Lab, who converted the directory to an Oracle database and provided a simple forms-based interface with remote "telnet" and "ftp" access. In late 1993, however, outside project funding ran out, bringing data development to a halt (and slowing the Council's other activities) until late 1995, when a new grant from the U.S. Environmental Protection Agency revived the Council's activities and its interest in shared information systems. In the nearly two year hiatus, the EDIMS Coordinator and a handful of students at the University of New Hampshire built a simple World Wide Web interface to EDIMS, which gradually attained greater visibility as state and provincial agencies gained Internet access. However, after being "dormant" for such a long and crucial time, EDIMS (and the Council) seem to have had difficulty picking up where they left off: many agencies had made decisions on computing and networking without EDIMS' input, and their concern for regional Gulf of Maine issues had waned. As a result, in mid-1996, participation in the 15-member EDIMS Committee was low (attendance overlap between consecutive meetings was small; people didn't know each other's names), coordination between American and Canadian participants was limited, and funding was sufficient for little more than maintenance. In early to mid-1996, EDIMS seemed to be struggling to define its role in a decentralized, complex technical and institutional environment.

The following sections present a bit more detail on the context EDIMS has fit into, the choices that have shaped it, and the measures of size and quality it has attained.

3. Context of the infrastructure

To understand EDIMS, it's important to know its institutional and technical surroundings, both past and present. Indeed, the Committee that oversees EDIMS interacts with other parts of the

Gulf of Maine Council, and with the region's states, provinces, and oceanographic research centers. Furthermore, EDIMS is one of several users and providers of ocean-related information, with which it maintains varying degrees of collaboration.

a. Institutional context

The Gulf of Maine Council on the Marine Environment is a coalition of state and provincial governments: Massachusetts, New Hampshire, Maine, New Brunswick, and Nova Scotia. Federal agencies in both countries, while not formally Council members, play an influential observer role on many of its committees, and provide much of its funding (state and province contributions have been mostly in-kind).

The Council's role in state and provincial governments is only advisory, rather than legislative or executive: it "encourages" its partner agencies to carry out the various elements of its Action Plan. According to the Canadian co-chair of the Committee that oversees EDIMS, this informal relationship lets Council committees define shared initiatives without much administrative baggage or political opposition. However, it seems to have limited the Council's influence on the agencies it's supposed to represent, and it may have kept the Council's Committees overly dependent on the strengths and interests of their individual participants (rather than founded on a clear charter shared and supported by partner agencies).

The Committee that oversees EDIMS interacts regularly with the Council's other Committees. In particular, it has worked for several years with the Council's Monitoring Committee to publish coastal ecosystem indicators online. Also, wider public access to the Web in recent months has prompted interactions with the Council's Public Participation Committee.

EDIMS also has relationships with several major research centers in the Gulf of Maine region, including the Bedford Institute of Oceanography in Nova Scotia and its American counterpart, the Woods Hole Oceanographic Institution in Massachusetts. The Council is also affiliated with two groups of researchers on the Gulf of Maine, the Gulf of Maine Regional Marine Research Board and the Regional Association for Research on the Gulf of Maine. Their relation to EDIMS is strengthened by the participation of the University of New Hampshire's Ocean Process Analysis Lab in both groups.

Other groups related to EDIMS include the Canadian Atlantic Coastal Zone Information Steering Committee (ACZISC), which coordinates the production and interchange of coastal and ocean data in the Canadian Atlantic provinces (Nova Scotia, New Brunswick, Prince Edward Island and Newfoundland and Labrador). ACZISC has a Database Directory Working Group, which maintains and regularly updates a directory of datasets on Atlantic Canada's coastal zone. The Canadian portion of the EDIMS dataset directory was drawn from this directory. The chair of ACZISC, and the chair of the Database Directory Working Group, both sit on the Committee that oversees EDIMS.

Finally, besides resource managers or researchers, several political associations are part of the EDIMS context, including the New England and Maritime Governor's Conference and the Council of Maritime Premiers. Their influence on EDIMS is indirect: their overall policy recommendations influence the Council's priorities and thus its demands on EDIMS.

b. Technological context

In addition to its institutional setting, EDIMS is tied to an increasingly complex technological context. For several years, one impediment to its growth was that most of its intended audience (resource managers and planners outside the university research setting) had no access to its information resources. Most state and provincial agencies in the region did not have Internet access, or adopted it in a hesitant, piecemeal fashion. (For instance, the first chair of the EDIMS Committee had to rely on a personal Internet mailbox for professional purposes.) The advent of Internet access in a majority of the region's agencies represented a boon for EDIMS, enabling its wider use and adoption among resource managers and state planners. On the other hand, now that these people have so many other sites on the Web to compare it with, they may demand a lot more from it. Already, members of the Public Participation Committee are demanding a "more interesting" front page for EDIMS.

Although Internet access is taking hold overall, the pattern varies a lot across the region. At one end of the spectrum, Nova Scotia's provincial agencies and Environment Canada have used the Internet for several years to communicate and share information. At the other end, some key agencies in New Hampshire remain hesitant or even unwilling to obtain basic Internet access. Major networking initiatives are underway in most of the states and provinces: highlights include Maine's NYNEX-funded network linking state agencies, schools, and libraries, Nova Scotia's province-wide fiber-optic network, and New Hampshire's ResourceNet, linking its state planning office and geographic data center.

The states and provinces each have a center for Geographic Information Systems (GIS) which archives digital information on land and natural resources and provides technical assistance to partner agencies. As of mid-1996, all of these GIS centers are looking into some form of online distribution of data and metadata. Their relationship to EDIMS ranges from close collaboration (the Committee's Canadian co-chair represents the Nova Scotia Land Use Committee, which coordinates all of Nova Scotia's GIS work) to mere acquaintance (New Hampshire's GIS center has had almost no interaction with EDIMS—despite being housed in the same building at the University of New Hampshire). Even though these GIS data centers emphasize land, not ocean, resources, most involved agree that their overlap with EDIMS would be fruitful.

4. Infrastructure choices

Given the goals set forth for EDIMS, and its institutional and technological context, it becomes possible to evaluate the choices—some organizational, others technical—made in its design and implementation.

a. Institutional arrangements

The most significant organizational choice in building EDIMS was probably the decision to host it at the Ocean Process Analysis Laboratory at the University of New Hampshire. This ensured easy access to advanced computing resources (such as a Unix server) and to the research community. The university setting also allowed EDIMS to survive an unfunded downtime: between late 1993 and mid-1995, University of New Hampshire staff and students not only kept it online, but turned it into a Web site and added a few information sources to it. However, a Committee member in

Nova Scotia suggested that low participation on the Canadian side may have been due to a perception of EDIMS as "a UNH research project, down there in New Hampshire."

b. Technological design

The first two years of the EDIMS Committee's existence were busy ones. It commissioned an in-depth survey of user needs, compiled a regional directory of coastal and marine datasets, and built the prototype telnet interface to this directory and ftp access to a few datasets. This directory was a centerpiece of the EDIMS design: it described data formats, dates, sizes, and distribution contact points. The Canadian portion of this directory was drawn from the database directory compiled by the Canadian Atlantic Coastal Zone Information Steering Committee (ACZISC).

Although the Council's intent for EDIMS was to link distributed environmental information resources through a "common regional protocol," the initial EDIMS prototype was housed on a single server at the University of New Hampshire. Through the database directory, EDIMS was to evolve from a centralized archive into a distributed system that would draw on data suppliers throughout the region. However, few keepers of related data rose to the task of putting their data online, so EDIMS has remained a central archive. The delay in moving to a distributed data model has caused some tension with strong proponents of decentralized systems. The Canadian provinces, in particular, had major concerns about creating a centralized database. These concerns may have arisen from their Land Records Information System, a regional initiative to create a centralized database some 20 years earlier. This was eventually abandoned in favor of a cooperative effort to create, coordinate, and maintain a center for "primary data" (geodetic control, basemaps, parcel maps), which split along provincial lines after a decade of data development due to a variety of technical and organizational difficulties. (As a single organization, it could not maintain the expertise needed to oversee all scientific and socio-economic land-related data for three provinces; furthermore, the provinces were reluctant to give up control over their information to this "outside" data center.) Some may have perceived EDIMS as another such centralized archiving effort, and felt that it should be distributed even in its prototype phase.

A third highlight of EDIMS' technical design was the choice of networking protocols — or lack of choice. The team at the University of New Hampshire advocated using the Internet Protocol (IP);

but other members of the Committee felt this would exclude government agencies, and maintained that other forms of networking (Banyan, modem) were needed. Some on the Committee would say that "the EDIMS concept was a bit ahead of its time; suddenly the Internet made it all possible." In late 1993, IP began its explosive expansion through every sector of business and government—but by then EDIMS had gone "dormant" for lack of funds. In essence, the EDIMS Committee seems to have missed the opportunity to build an advanced information service when finances and support were relatively plentiful (though not all on the Committee would agree that funding was ever adequate for the EDIMS goal); it now finds itself merely maintaining a service whose functionality (by the World Wide Web's rapidly rising standards) may seem limited.

5. Information sharing characteristics

The EDIMS effort can be evaluated in several ways. The following paragraphs outline several measures of the size and "quality" of EDIMS as an information sharing infrastructure, beginning with the kinds of information sharing it provides.

a. Forms of information sharing

EDIMS is primarily a repository and clearinghouse for coastal and ocean information, accessed through the Internet's World Wide Web protocol. The target audience is researchers and resource managers in the Gulf of Maine (though actual usage patterns may be quite different—see below).

A main feature of the EDIMS Web site is the directory of coastal and marine datasets, stored in a relational database and queried through a World Wide Web forms interface. Users can search on keywords, geographic coordinates, and date ranges. The EDIMS front page also points to several sets of data archived locally, including (i) sea-surface temperature maps (several images a day since mid-1993); (ii) hydrographic data from the Atlantic Fisheries and Adjustment Program (AFAP), from 1993; (iii) oceanographic indicators in the Massachusetts and Cape Cod bays, compiled in 1990-91; and other bathymetric, meteorological, and ecological datasets. EDIMS also supports the Council's administrative communication needs through several mailing lists, electronic-mail forums, and bulletin-board systems.

b. Size of the infrastructure

One measure of an information sharing infrastructure is the number of sites that provide information. In the case of EDIMS, most of the data is still local at present, but a few outside servers are being linked in via hypertext: the US Geological Survey and NOAA at the Woods Hole Oceanographic Institution in Massachusetts; Environment Canada; and others. Besides those who have put their own data online, several others have contributed data to be served up on the EDIMS server. These include the Massachusetts Dept. of Coastal Zone Management, the Bedford Institute of Oceanography, and the US Fish and Wildlife Service's Gulf of Maine Program.

Another measure is the volume of data available through EDIMS. Besides its hypertext pointers to agencies and providers in the region, the EDIMS server itself provides about 170Mb of data, including sea-surface temperature maps (101 Mb); (ii) Atlantic Fisheries and Adjustment Program (AFAP) data (25 Mb); and (iii) Massachusetts and Cape Cod oceanographic data (23 Mb).

A third measure is the traffic volume on the main EDIMS server, and the number of distinct sites using it. According to its logs, in the month of July, 1996, it transferred 22 Mb (3,300 files). (By comparison, the "hub" server for the Great Lakes Information Network transferred 51,000 files totaling 485 Mb in the same month.) EDIMS' traffic grew by about 19% per month from January to July, 1996. Usage figures for EDIMS, due to their small size, are quite sensitive both to individual users (a persistent user may add 10 or 15% to the month's usage in a single afternoon) and to automated Web search engines (which account for 15-25% of all data requests, more than any actual user). Top users (other than search engines) include the State of Maine and Canada's Department of Fisheries and Oceans, but also the Institute for Global Communications and the University of Washington— both well outside the "target" audience. 1996 has brought dramatic growth in the number of distinct sites accessing EDIMS, from just over 100 in January to nearly 800 in July. This is most likely due to the increased use of Web search engines, and suggests that most or all of EDIMS' growth in traffic may be outside its intended audience.

In summary, EDIMS is still an infrastructure of modest size, growing at an average rate, with an increasingly diverse user group. Although it is being used by researchers and resource managers in the Gulf of Maine, their use has grown only slowly, and is outweighed by that of outside users.

c. Quality of the shared information

Having provided several measures of the size of this infrastructure, we now turn to an assessment of its "quality." This is in part a property of the information it provides to users: its precision, accuracy, timeliness, and usability.

i. Precision and accuracy

As mentioned earlier, one of the key features of the EDIMS server is the Gulf of Maine Dataset Directory. EDIMS Committee members generally consider this a useful resource; but it was compiled in 1991-1992 and has not been updated since. Many of these listings, especially contact names and telephone numbers, are probably out of date. Perhaps for this reason, the directory sees only light use. (A comprehensive directory update was planned beginning in mid-1996.)

Besides this directory, the actual datasets on EDIMS have only seen light use as well, so their precision or accuracy have not been a great concern so far. Nonetheless, each set of data on the EDIMS server is accompanied by a free-text description of survey methods, time frame, data formats, and often a published report that analyzes and documents the dataset. This should at least help users decide whether the precision and accuracy of the data are adequate for their purpose.

ii. Timeliness and concurrency

These criteria measure how well the data's update frequency matches that of the phenomenon being measured and that of the user's analysis or decisions. These criteria have only limited applicability to EDIMS, given that most of its data are fixed measurements made at a specific date and time. One exception: the Sea Surface Temperature maps, which the EDIMS coordinator at the University of New Hampshire transfers from a NOAA server every morning. (Interestingly, one of the few instances of EDIMS user feedback came from a group of tuna fishermen, who sent electronic mail requesting more frequent updates to the sea surface temperature maps.) Another exception is a set of data on New Brunswick river flows, which until March 1996 was updated daily from a server in Saint John, NB. Here again, because of the light use of these data, their timeliness or concurrency have not been a prime concern.

iii. Usability and encapsulation

Another measure of EDIMS information: the amount of knowledge and tools needed to interpret and use it. Obviously, EDIMS' ease of use varies among its uses and users. Administrative information, such as e-mail forums and address lists, is easily accessed through a Web browser such as Netscape. The dataset directory is easy to search through the forms interface, given some knowledge of geography and either oceanography, fisheries, or meteorology. The archived datasets are a different story: they are provided as ASCII tables and files whose correct use requires tools and expertise beyond the reach of the typical practitioner.

In its planning for distributed data, EDIMS is looking into use of a protocol devised by the international Joint Global Ocean Flux Study (JGOFS). This protocol would provide transparent access to distributed data in a variety of formats—a key ingredient of the EDIMS design. However, the JGOFS user interfaces, built to work with modeling software, are still quite complex: the preferred client among JGOFS implementers is Mathworks' MatLab function-analysis software. Another candidate data protocol, the Distributed Ocean Data System, under development at the University of Rhode Island, has an expanded set of user-interface options, though still geared to scientists. In general, fitting a widely usable front end onto a distributed data protocol will be key to EDIMS' data architecture, and to its wider acceptance as a useful, reliable data clearinghouse.

d. Quality of the information infrastructure

In addition to information characteristics, the quality of an information sharing infrastructure depends on aspects of the sharing mechanism itself: its flexibility for two-way interchange, its ability to grow in size, and its ease of integration with diverse local data systems.

i. Reciprocity

This aspect of the infrastructure describes how easily information receivers can also be providers. In principle, this is easy to achieve: to include new data providers, just add hypertext links to their server at the EDIMS "hub." In practice, though, maintaining a server with dynamic operational or scientific data (i.e. not just electronic "brochures") requires expertise that most EDIMS participants currently lack. EDIMS does provide a mailing-list server and a Web-based bulletin-board system, which enable two-way, free-text communication among participants. These are helping to establish EDIMS as a focal point for the Council's (and, more generally, the Gulf of Maine's) regional communications and data development. But overall, EDIMS is still far from its goal of linking distributed data throughout the region. Towards that end, EDIMS staff have been looking into a variety of means (including JGOFS, as mentioned above) to link in distributed data sources around the region; as these sources begin to put data online, EDIMS' Web-based design should be easily able to link in their contributions.

ii. Scalability

EDIMS' Web-based design also makes it easily scaleable: if data sources are maintained on autonomous Web servers, it's easy to provide hypertext links to them. But again, few data sources are online: most data are sent in their native format to data coordinator Karen Garrison, who manually installs them on the EDIMS web server. This method is still manageable given the small number of contributors and data sets, but is by no means scaleable: the contributors, datasets, and datatypes should grow from the present handful to dozens or hundreds, installing and updating data on the EDIMS server would become a certain bottleneck.

iii. Non-intrusiveness

To date, this aspect of the infrastructure has not been a prime concern, because few of its participants have considered how to link EDIMS to their work or their own data systems. Interestingly, the EDIMS design contains little or no discussion of standards; current designs have favored setting as few standards as possible. In particular, the EDIMS committee has only recently begun to evaluate metadata standards and reconciling the Canadian and U.S. standards. Once data sources start to come on-line, a shared data protocol such as JGOFS may prove useful as a "minimal" standard, letting data maintainers keep their information in whatever form is most useful to them locally, while still allowing outside access to it in a consistent fashion.

6. Infrastructure impacts

Perhaps the most important assessment of an information infrastructure is its impact on people's work: their efficiency, effectiveness, or satisfaction.

To date, EDIMS has had few tangible impacts on resource management or policy decisions. The Council's policy decisions are still generally independent of EDIMS information; and the individual states and provinces continue to manage coastal and marine resources within their political boundaries, with little attention to regional impacts or trends, and with minimal use of the information or communications provided by EDIMS.

Nonetheless, several participants in the EDIMS Committee report that it has facilitated informal communication, cooperation, and trust between organizations. Others praise EDIMS as a cheap and effective replacement for telephones or mass mailings. So, although its tangible impacts to date may be minor, EDIMS may be laying the human and organizational groundwork for future information-based collaboration.

In particular, under its revised five-year Action Plan (1996-2001), the Council is calling for better public access to ecological data, for consolidated information on toxic contaminants, and for improved communication among research and monitoring programs. These elements in the Action Plan suggest that the Council is beginning to incorporate regional environmental data into its objectives. They also indicate an emerging reliance on EDIMS as a central tool in working towards those objectives. As the Council's priorities become more information-intensive, EDIMS may yet get to play its intended pivotal role in guiding environmental decisions in the Gulf of Maine.

7. Challenges and lessons

As detailed earlier, EDIMS was expected to be further along than it is in coordinating information resources and supporting regional decision-making in the Gulf of Maine. EDIMS participants attribute its slow development and acceptance to a combination of four factors: cultural difficulties (researchers are slow to structure and release their data; state personnel neglect to read their e-mail), organizational inertia (the Council's thinking and decision-making make little use of physical data), technical hurdles (there's still no easy interface to the data sets), and financial constraints (current funding allows for little more than maintenance).

In addition to these stated barriers, several others seem to have had an additional impact. First, there seems to be no clear shared goal or sense of inter-dependence among EDIMS partner agencies. Even though the different parts of the Gulf of Maine are ecologically inter-dependent, this physical and biological interdependence has thus far had little effect on the policies that the Council is concerned with (the well-known Gulf of Maine fisheries crisis is outside its purview). So, despite the Council's efforts, and EDIMS' relationship-building, the states and provinces are still largely focused on the natural resources and land information that fall within their own political boundaries.

Another apparent impediment to EDIMS has been its relationship with its participating agencies. The EDIMS Committee carries little weight as a coordinating body: rather, it seems to be a loose group of individuals who share a personal interest in the Gulf of Maine. The influence of these individuals on their home agencies is limited; and these agencies' contributions to the Council and to EDIMS are usually only in-kind (small percentages of staff time; travel to meetings) rather than actual funds to enable new initiatives.

The Council's relationship to EDIMS has itself seemed less than effective at times. For instance, in the first two years, it was unable to provide a clear set of goals for EDIMS' early information-building efforts, and finally gave its formal approval to EDIMS in late 1993, just as funding for any real implementation was running out.

So it seems EDIMS was largely hampered by a lack of clear and timely goals, combined with its weak, informal relationships with both partner and parent organizations. Faced with these hurdles, EDIMS was restricted to low-impact, minimal-cost activities. The Council's stated focus, in its new revised Action Plan, is to promote the protection and restoration of coastal and marine habitat. Will this be a sufficiently clear and compelling shared goal to elicit real commitments (of money, time, and expertise) on the part of the states and provinces? It's too early to tell. One thing is clear, however: building EDIMS into the high-quality information infrastructure it was envisioned as will be costly and difficult, from both a technological and organizational standpoint.

8. Conclusions

The Gulf of Maine Environmental Data and Information System (EDIMS) is an ambitious attempt to bring together the information resources of a large and diverse region. It made rapid progress in the first two years of its existence, and built an early foundation for a potentially high-quality information infrastructure. However, its progress has slowed since then, ostensibly for lack of funds, but also for lack of a common goal and a clear networking strategy. It has also suffered from its weak influence on its participating agencies, and a lack of direction from its parent agency, the Gulf of Maine Council on the Marine Environment.

Currently, its chief contribution to the Gulf of Maine community is a communications medium; as such, it is starting to gain wider acceptance as more and more agencies gain Internet access. This may win it the institutional support it needs from agencies throughout the region—and from the Council itself. Through the team at the University of New Hampshire, EDIMS has also built strong ties to the Gulf of Maine research community, from which it will get advanced technical support. EDIMS' experience suggests that both kinds of support will be greatly needed for the task of building a distributed regional information infrastructure.

Chapter 6: The Northwest Environmental Database, Coordinated Information System, and StreamNet

1. Overview

For over a decade, the U.S. Department of Energy's Bonneville Power Administration has worked with state, tribal, and federal agencies to harmonize fisheries, wildlife, and hydroelectric facilities in the Pacific Northwest region. This region is linked geographically and economically by the Columbia River basin, which covers an area roughly the size of France (Figure 1).

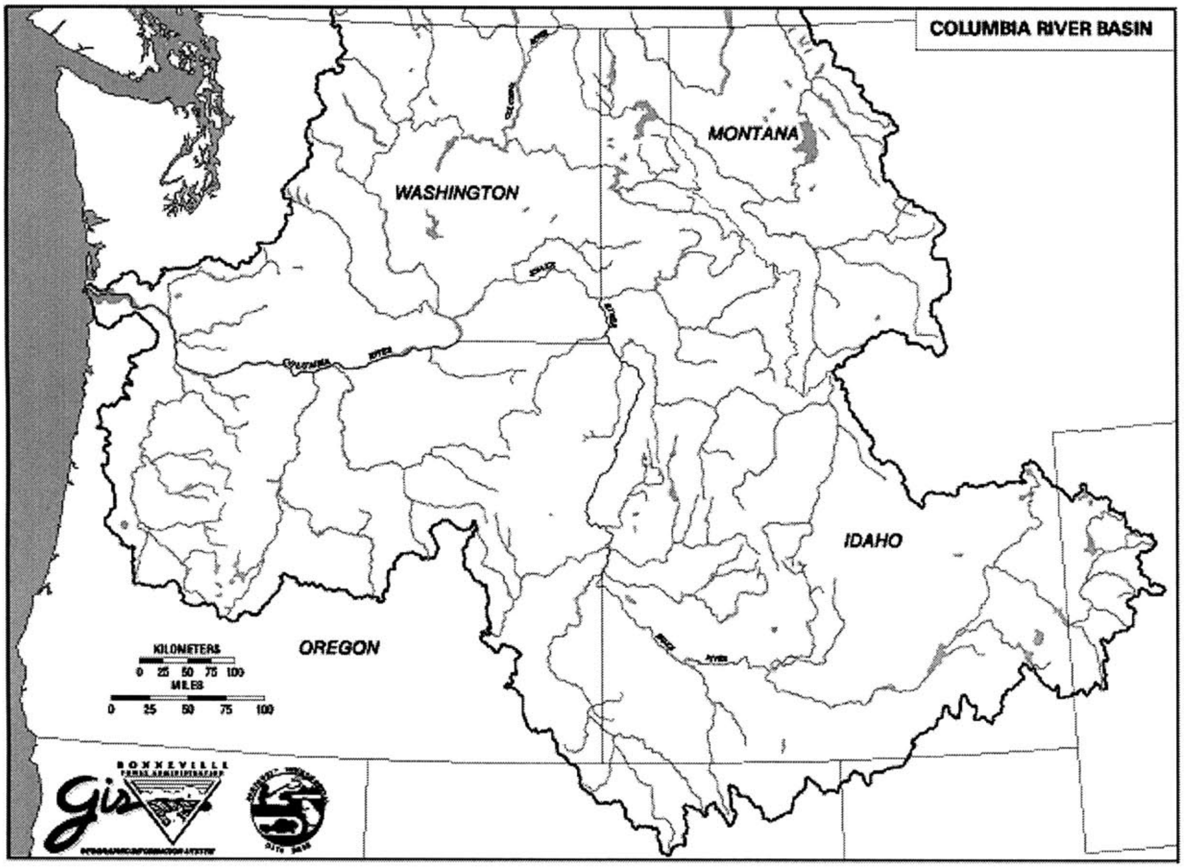


Figure 6-1. Pacific Northwest States and Columbia River Basin

To support these activities, these agencies built two collections of digital data on stream-related fisheries and wildlife, the Northwest Environmental Database (NED) and the Coordinated Information System (CIS). NED was a collection of geographic datasets on fisheries, wildlife, cultural resources, dams, and other features throughout the Pacific Northwest. CIS was a tabular database focused on anadromous fisheries (i.e. ocean-migrating species such as salmon and steelhead) within the Columbia River basin.

Data for both systems came from many different sources and were shared via state and tribal coordinators using a common geographic reference, the River Reach File identifiers devised by the U.S. Environmental Protection Agency (EPA) and the U.S. Geological Survey. Both systems were funded by the Bonneville Power Administration and guided by the Northwest Power Planning Council, a regional agency representing the four Northwest states. Yet despite their similarities, NED and CIS were conceived and built separately: the former to plan hydropower development and protect fish and wildlife; the latter to monitor the progress of anadromous fisheries programs. They took quite different approaches to building a regional information sharing infrastructure. NED consisted of topologically complete geographic data residing in and maintained by individual state agencies, loosely integrated by means of the River Reach identifiers and a set of cross-reference tables. CIS followed a more traditional approach, with centralized, tabular data maintained and disseminated by a regional consortium, the Pacific States Marine Fisheries Commission.

Both systems facilitated complex collaborative environmental management in the region. NED supported habitat assessment for endangered species, hydroelectric permitting, and a variety of other activities; CIS supported the management, and rehabilitation of salmon and steelhead species in the Columbia River and its tributaries, several of which have become severely threatened in recent years.

After several years of parallel existence, NED and CIS were merged in early 1996 into a regional "aquatic resource information network," dubbed StreamNet. The initial focus of this new information system is the assessment and protection of anadromous fisheries; ultimately, its founders hope to build it into a general-purpose information resource. Meanwhile, as the salmon crisis and other environmental controversies unfold in the Pacific Northwest, dozens of private, state, federal, and tribal groups continue trying to protect and restore a valuable ecosystem permanently altered by human activity. StreamNet may have a pivotal role to play in these efforts by maintaining regional consistency, facilitating consensus, and infusing controversial decisions and complex research with the best possible environmental information.

StreamNet and its predecessors offer useful lessons for the design of modern information-sharing infrastructures because of their decentralized data storage and control, their relative longevity, and the clear impacts they have had on tasks and processes of environmental management in the region. This study sketches the contexts in which these systems were built, the choices made in their design, the impacts they achieved, and the challenges they encountered, in order to draw general lessons on the sharing of geographic information.

2. Brief history

a. The Northwest Power Act

In December 1980, the U.S. Congress passed the Northwest Power Act (U.S. Congress, 1980), calling for electric-power conservation and planning in the Pacific Northwest region. Recognizing the adverse ecological impacts of hydroelectric power, the Act also included measures to "protect, mitigate, and enhance the fish and wildlife, including related spawning grounds and habitat, of the Columbia River and its tributaries" (§4(h)(1)(A)). The Act created a regional inter-state compact, the Northwest Power Planning Council, with a mandate to draw up a long-term Electric Power Plan for the four-state region, and a Fish and Wildlife Program for the Columbia River basin; and to oversee their implementation with funds from the Bonneville Power Administration. The Northwest Power Act was intended as a compromise between utilities' energy needs and the public's call for energy conservation environmental protection. However, beginning in the early 1980s, the demand for power dropped, and the Northwest Power Act ended up being "the blueprint for a laboratory of energy and environmental conservation" in the region (National Research Council, 1995).

In its call for a regional Fish and Wildlife Program, the Northwest Power Act spelled out three requirements in particular:

> **§4(h)(6).** The Council shall include in the program measures which (...) will
>
> **§4(h)(6)(A).** complement the existing and future activities of the Federal and the region's State fish and wildlife agencies and appropriate Indian tribes;
>
> **§4(h)(6)(B).** be based on, and supported by, the best available scientific knowledge; (...)

> **§4(h)(1)(A).** The program shall, to the extent possible, be designed to deal with [the Columbia] river and its tributaries as a system.

Accordingly, Bonneville and the Power Planning Council began the Fish and Wildlife Program with a regional, systematic survey of stream-related natural resources, the Hydro Assessment Study. This was conducted in two parts, with quite different methods and outcomes: a detailed study of anadromous fisheries and Indian cultural values, and the Pacific Northwest Rivers Study, which examined everything else, as follows.

b. Pacific Northwest Rivers Study and Northwest Environmental Database

For the Pacific Northwest Rivers Study, Bonneville contracted with the states of Idaho, Montana, Oregon, and Washington, together with the region's Native American tribes and Federal land management agencies, to co-write assessment guidelines; these agencies then spent over two years (1984-86) collecting a broad range of data on the region's rivers and streams: fisheries, wildlife, botanical and geological features, recreation, historic and archeological sites, and legal constraints. The data were qualitative in nature, describing a stream's significance (high / medium / low) for each of the resource categories, as defined by panels of local specialists in each area. In keeping with the Electric Power Plan, the study's scope went beyond the Columbia River basin, to encompass the entire four-state region.

Midway through the Rivers Study, Bonneville convened a "Geographic Information System Task Force" to design and plan for an information system which would integrate the data from the four state Rivers Studies, including anadromous fisheries, for a variety of purposes. This data system consisted of digital geographic "layers" of stream-quality rankings tied to spatial features through the Environmental Protection Agency's nascent River Reach Files (BPA, 1985) at a scale of 1:250,000. Bonneville and the Council later added hydroelectric site information, and dubbed the resulting collection of data layers the Northwest Environmental Database, or NED.

NED was envisioned as an evolving set of databases maintained in the four states, with summary information transmitted biannually back to Bonneville. It was also seen as geographic, rather than tabular, in nature: "Digital spatial analysis is a desire of both [Bonneville] and the Council," reported the GIS Task Force in 1985. For the most part, this vision came true: each of the four states distributed its portion of NED as a microcomputer-based "Rivers Information System," and all four states integrated these data to varying degrees into their own geographic data collections. Bonneville would extract summaries from the state databases, and incorporate them into a region-wide set of GIS data layers, while maintaining its own database of hydroelectric sites obtained from the US Army Corps of Engineers.

Thanks to the Northwest Rivers Study and the parallel anadromous fish study (detailed below), the Council was able to amend its Fish and Wildlife Program in 1988, declaring 44,000 miles of streams "Protected Areas," off-limits to all hydroelectric development. An excerpt from the Amendment shows how much it relied on the Rivers Study information:

> Beginning in 1983, the Council directed extensive studies of existing habitat and has analyzed alternative means of protection. (...) The Council, relying on these studies, designated certain river reaches in the Basin as "protected areas," where the Council believes hydroelectric development would have unacceptable risk of loss to fish and wildlife species of concern, their productive capacity, or their habitat. (NWPPC, 1988)

The Council's Protected Areas amendment, combined with an unexpectedly low energy demand through most of the 1980s, brought hydroelectric development almost to a standstill in the region. As a result, after 1988, NED saw little use for system-wide facility siting or regional power planning. Instead, it served a variety of other purposes, such as assessing species health and habitat quality, or preparing maps and guides. Meanwhile, as NED's state partners became increasingly focused on site-specific environmental assessment tasks, they found NED's qualitative measures and its coarse-grained geographic scale increasingly ill-suited to their needs. Thus, faced with competing priorities, many of these agencies allowed their portion of NED to fall out of date, and out of consistency with other NED components in the region.

c. Anadromous fisheries and the Coordinated Information System.

As the Northwest Rivers Study was underway, in a separate part of the Hydro Assessment Study, the Northwest Power Planning Council was collecting detailed, quantitative data on anadromous fisheries. For several years, beginning in 1983, the Columbia River Inter-Tribal Fish Commission (CRITFC), along with Bonneville, the Council, and other state, federal, and tribal organizations, debated how to coordinate, maintain, and disseminate the resulting anadromous fisheries data. This was a slow process: Bonneville first designated funds for a Coordinated Information System in 1988; actual databases began to be constructed in 1992; and the Pacific States Marine Fisheries Commission assumed responsibility for the project in 1993. CIS grew to be a large, complex set of data compiled from many published and unpublished sources. State, federal, and tribal fisheries managers would send information to the Pacific States Marine Fisheries Commission, which distributed it periodically to users via diskettes, reports, and an electronic bulletin board.

CIS was funded by Bonneville and built on the same unit of analysis as NED (the stream reach)—closely related to and in many ways an offshoot of NED. "I view CIS as conceived out of "NED envy,'" recalls one NED participant, "I have always seen them as two parts of the same puzzle. [Bonneville was] hornswoggled from day one to pay twice for the same work!" Yet CIS's proponents emphasized its detailed, quantitative, purely tabular information, specifically focused on anadromous fisheries (distributions, habitat, life stages, and genetics): "it's quantitative data, which means you can work with it." Moreover, as one member of the CIS steering committee pointed out, "We don't do analysis here": CIS emphasized raw numbers as more "neutral" data than NED's qualitative judgments, more acceptable to all parties involved. One final difference: in keeping with the Fish and Wildlife portions of the Northwest Power Act, CIS data covered only the Idaho, Oregon, and Washington portions of the Columbia River basin (Montana has no anadromous fish).

Thus, no-one saw the two studies, or the data they produced, as co-equal. Many considered anadromous fish to be the true focus of the Hydro Assessment Study, with geographic references and additional data categories added in on a secondary basis. Others felt that the anadromous fish study should have been included in the Rivers Study's comprehensive, distributed, geographical approach. The two views are summarized by the following exchange, overheard at a Council meeting: "CIS is the tree, and NED is a branch on that tree. –No, NED is the tree, and CIS is the branch." This difference of views persisted even as Bonneville was merging the two programs (into StreamNet) in 1996: maintainers of CIS simply renamed their stand-alone database "StreamNet (formerly CIS)," and put a hypertext link to it on the StreamNet website, alongside interactive map catalogs and a wide range of stream-related links.

3. Context of the infrastructure

The context of these two data sharing efforts helps to explain the choices made in their design and implementation; changes in their context help to explain why some approaches may have been difficult to sustain.

a. Institutional context

i. Many conflicting stakeholders

Both NED and CIS arose in a context of antagonism between developers, protectors, and competing users of natural resources. As Dietrich (1995), Lee (1993), and others have detailed, the institutions concerned with the Columbia River Basin form a complex, overlapping web of regional, state, federal, tribal, and private groups, all with different, often conflicting objectives and views of the river. In its recent "Salmon Governance" report to Congress, the Council referred to "a constellation of agencies, courts and other entities" that have shaped the development and management of the Columbia River (Northwest Power Planning Council, 1996).

Nearly all inter-agency cooperation in the Columbia River region has been mandated by court orders or federal legislation (e.g. the Pacific Salmon Treaty, the Salmon and Steelhead Conservation Act), and often facilitated by improved data and analyses that helped competing users to agree on basic assumptions. Yet despite federal legislation and the promise of improved decision-making information, cooperative projects such as NED and CIS often had to adopt complex arrangements to work around sharp disagreements between agencies. For instance, as one CIS participant recalls, Idaho's Fish and Game Department and the Shoshone-Bannock Tribes would not contract with the Columbia River Inter-Tribal Fisheries Commission due to ongoing court battles; so the contract for their CIS work was made through a third party, the Columbia River Fish and Wildlife Authority.

ii. Other regional data sharing efforts: reinforcement and competition

When NED and CIS first got underway in the mid-1980s, many inter-agency efforts had been formed over the years ("the Columbia River Basin is the capital of inter-agency fishery committees," one long-time observer told me); but paradoxically, information sharing between agencies had remained rare. The Northwest Power Act and its resulting Hydro Assessment Study were seen as a pioneering experiment in inter-state and state-federal cooperation: one Council member considers that one of the greatest accomplishments of the Northwest Rivers Study was simply persuading states and tribes to release their data outside their jurisdiction. In subsequent years, the region saw several other efforts in regional cooperation and data sharing. These encouraged agencies to coordinate their information and activities and expanded the channels of communication among them—while also complicating the choices among competing standards and affiliations.

From the mid-1980s to the early 1990s, Bonneville took part in the Northwest Land Information Systems Network (NWLISN), an interagency committee that built a geographic data clearinghouse for Washington and Oregon, the Spatial Data indeX (SDX), maintained by a research group at Portland State University (Dueker and Vrana, 1995). Despite an enthusiastic start, the Network's informal structure proved hard to sustain based only on a shared need, in the absence of dedicated funds or signatory commitments. In the words of one NWLISN participant, "There was no real meat involved: we just passed the hat whenever data needed to be collected." Nonetheless, in 1986 the Network led the US Geological Survey to begin upgrading the River Reach files to the 1:100 000 scale in the four-state region (Fisher, 1993), a move which finally came to fruition, largely through NED, in the mid-1990s as the four states finalized the Reach Files and upgraded their rivers data to this level of detail.

Also at the federal level, President Clinton's 1992 Northwest Forest Plan (EOP, 1993) responded to a long-standing controversy over old-growth forests by calling for regional ecosystem management in the US Forest Service and the Bureau of Land Management. As detailed by Yaffee (1994), this forced the Forest Service in particular to undergo deep changes in its goals and *modus operandi.* In response, these two agencies undertook major efforts in data collection and standardization efforts, and in sharing information among their own units and with outside agencies (including several NED and CIS participants). The first of these programs, known as the Inter-organization Resource Information Coordinating Council (IRICC), covered western portions of Washington, Oregon, and northern California, with a strong initial focus on the range of the northern spotted owl. A second program, known as the Eastside Assessment, studied forests in eastern Washington, Oregon, and Idaho. A third program, known as PACFISH, focused on Pacific salmon and steelhead habitat from California to Alaska. These programs brought together many resource data managers in the Pacific Northwest region; they funded some NED-related data development (the Eastside Assessment helped fund a 1994 update of the Idaho Rivers Information System); they led to detailed data standards (such as PACFISH stream habitat criteria), and contributed detailed information (the present-day StreamNet credits IRICC for its aquatic habitat data).

iii. Increasingly mature & formal state data sharing

Over the course of NED's and CIS' existence, each of the region's four states independently defined increasingly detailed standards and procedures for data exchange. Each defined a "lead" agency for GIS, a Geographic Information Council, and written guidelines for inter-agency cooperation and data exchange. Under funding from IRICC, Oregon and Washington made formal agreements (known, respectively, as "Data '96" and "Baseline '97") with state and federal agencies for long-term cooperative base data development. Thus, regional data sharing efforts such as NED or CIS had to take into account increasingly complex underlying institutions and relationships in their contracting and funding.

iv. Informal communities

In addition to formal guidelines and institutional agreements, increasing use of computers and geographic information systems led each of the states to set up technical working groups to facilitate communication and coordination. One regional forum among state GIS specialists came about unexpectedly: when the U.S. Geological Survey released the new 1:100,000 River Reach files in 1990, state agencies found them riddled with errors; a "user's revolt" ensued (in the words of one state GIS manager), and Bonneville contracted with the states to finish the job, which took until mid-1995. In so doing, the states traded information and learned from each other in the area of data management and GIS. These informal technical relationships were further strengthened by mechanisms such as the annual Northwest Arc/Info Conference. As a result, state GIS specialists in Idaho, for instance, know far more about their counterparts in Montana or Oregon than those in neighboring Nevada, Utah, or Wyoming.

b. Technological context

i. Widely varying degrees of experience

In the early 1980s, as the statewide Rivers Studies got underway, geographic information systems (GIS) were beginning to shape the work of a few agencies in the region, including Bonneville and the US Geological Survey. In assembling information from the Rivers Studies, however, Bonneville's approach had to be very flexible, not only because of conflicts and mistrust between organizations, as noted above, but also because expertise varied widely across the region's federal, state, and tribal organizations.

Many agencies had little or no experience with digital mapping or data management: in 1985, for instance, Bonneville's GIS Task Force doubted that there was enough GIS capacity within the states or federal agencies to attempt cooperative digitizing of paper maps (BPA, 1985). End-user computing was also quite new to some agencies: the Northwest Rivers Study supplied Idaho's Department of Fish and Game with its first personal computer. Thus, NED and CIS couldn't count on a lot of technical input from Bonneville's partners; part of their work consisted of equipping state and tribal agencies with hardware, software, and training.

On the other hand, a few state agencies had years of experience with GIS and databases. Washington's Department of Natural Resources and Department of Wildlife were customers #3 and #19 of the Environmental Systems Research Institute (ESRI), maker of Arc/Info GIS software (now the leading "industrial-strength" GIS system). Montana's Department of Fish, Wildlife, and Parks began rating its stream resources systematically in the 1950s, and had maintained an extensive Interagency Stream Fisheries Database, initially on mainframe punch-cards, since 1979. Not surprisingly, these agencies were sometimes reluctant to conform to NED's specifications, which they saw as "a step down" from their own sophisticated data systems: "what they don't get from us in consistency, they get in information," said one NED participant. Even as GIS use increased throughout the region, these wide differences in expertise persisted and continued to complicate any coordination.

ii. Increasing prominence of GIS standards and networking

By the mid-1990s, all four states had built up their geographic data collections significantly, with large datasets at scales of 1:100,000 or 1:24,000; and each state's geographic information council had begun to set standards for information exchange within the state. Given their increasing sophistication, states became less likely to worry about conforming with NED in their data management decisions, given its coarse scale (1:250,000) and the "archaic" form of each state's Rivers Information System (generally a stand-alone database with a DOS menu interface). In most cases, the states incorporated NED's rivers data into their state GIS collections, modifying them with little concern for regional consistency. Their data management choices mostly followed their own needs in-state, with some adjustments to suit regional projects funded by the Forest Service and Bureau of Land Management: for instance, in 1993 IRICC's GIS team recommended a standard map scale of 1:24,000.

Furthermore, although the design and implementation of NED and CIS did not depend on a physical data network, several of their participating agencies began to explore the use of an increasingly ubiquitous Internet in sharing natural resource information. By the mid-1990s, some had come to rely on wide-area networks for their basic functions: Washington's Department of Natural Resources used the Internet to share data volumes with the Department of Fish and Wildlife across town, and to link its own 14 district offices across the state to large shared computing resources. Montana's State Library became one of the first nodes on the National Geospatial Data Clearinghouse, and led the move to build an Internet "backbone" linking state offices in Montana's capital.

Thus, regional data initiatives such as NED or CIS had to take into account increasingly complex data systems, rather than just building them from the ground up as they could in the early 1980s. In 1995, NED and CIS managers began to experiment with Internet data distribution, which led them to design StreamNet as a fully Internet-based information service, with distributed databases and maps based on the World Wide Web.

4. Infrastructure choices

In response to their context, NED and CIS made distinct choices in building an information sharing infrastructure, including both institutional arrangements (responsibilities and decision-making powers) and technical design choices (degrees of decentralization, standards, and geographic scale).

a. Institutional arrangements

Both NED and CIS were funded by Bonneville and guided by the Northwest Power Planning Council, but they adopted quite different organizational structures.

i. Northwest Environmental Database: stakeholder cooperation and interdependence

NED's institutional relationships were designed to involve a wide range of interested parties while keeping a fairly simple structure. In cooperation with the Northwest Power Planning Council, Bonneville funded state agencies in Montana, Idaho, Washington, and Oregon to collect and then maintain statewide rivers data. These counterparts had some discretion as to the data format, pace, and even scope of their portion of the study, and coordinated state, federal, and tribal sources of resource information within each state. NED brought together, in some cases for the first time, information and experts from quite different areas, with quite fruitful results: as one participant recalls, Montana's Department of Fish, Wildlife, and Parks provided initial data-management concepts in 1985; Oregon's Department of Energy then wrote up a data management plan and built a DB2 interface; the Washington State Energy Office proposed tying data to stream reaches; Idaho's Department of Fish and Game added color to Oregon's data interface; and thus NED gradually took shape. Several years later, Idaho's and Oregon's Water Resources Departments led the move to complete the 1:100,000-scale River Reach files. Data maintenance was entrusted to the state energy departments and fish and wildlife agencies, with regional coordination through quarterly reports and biannual meetings. Thus, to enlist the support of the major stakeholders, NED used interdependence and cooperation between federal and state government agencies, and between energy production and environmental protection.

ii. Coordinated Information System: neutrality through isolation from stakeholders

CIS' organizational relationships were a bit more complex. The Columbia River Inter-Tribal Fish Commission (CRITFC) developed the concept in 1984 and built institutional partnerships over several years. CIS began as a forum for exchanging salmon data between policy-level fisheries specialists; but it gradually turned into a database management project, and in 1993 CRITFC transferred its technical and administrative responsibility to the Pacific States Marine Fisheries Commission (PSMFC), which had several years of experience with managing large fisheries databases (PSMFC, 1996). CIS built institutional support through bimonthly meetings of its Steering Committee, which represented Bonneville, the Council, CRITFC, state and tribal fish and wildlife agencies, and the National Marine Fisheries Service, which played an important role after 1992 under the Endangered Species Act. Thus, to encourage support from all sides of the salmon controversy, CIS positioned itself as a repository of "uninterpreted," "neutral" data on anadromous fisheries, housed within a well-respected inter-state Commission with no regulatory or management authority.

The new StreamNet embodies both of these approaches, contradictory though they may be (one NED coordinator described it as "one of the continuing thorns in my side"). To emphasize StreamNet's independence from the region's stakeholders, a new Internet domain was created for it, *streamnet.org*, whose Web-based information services are developed largely on Bonneville computers but reside on PSMFC computers.

b. Technological design

Here again, NED and CIS represent quite different choices. NED's focus was broad, qualitative data, carefully tied to geographic data features; whereas CIS emphasized quantitative detail and bibliographic references, with only implicit links to geographic features.

i. Northwest Environmental Database: decentralized, qualitative, geographic

Initially, NED was not intended as a distributed database; but because the Council's efforts were to complement the region's existing fish and wildlife activities, NED was designed to strengthen, expand, and integrate, but not replace, existing data systems. This was accomplished through a distributed database model. This was an ambitious choice for 1985: Bonneville and the Council let each state determine most data management specifics, and in order to integrate and compare between states, they spent about a year building cross-reference tables between each state's river coding systems and the EPA Reach File identifiers. NED also emphasized spatial analysis and a comprehensive view of the ecosystem through use of GIS. To obtain a regional overview quickly, Bonneville and the Council opted for a geographic scale of 1:250,000, and chose to record only qualitative judgments (such as good / fair / poor habitat) from panels of biologists, rather than quantitative data (such as population counts for particular species).

ii. Coordinated Information System: centralized, quantitative, tabular

CIS took a more centralized approach than NED: the Pacific States Marine Fisheries Commission (PSMFC) managed a tightly integrated set of data tables, for which it collected data mostly by traditional methods such as reading printed reports (though some contributors did begin to shift to electronic data submissions). River Reach numbers, at a scale of 1:250,000, served as identifiers, but their geographic role in CIS' tabular data was only implicit. Although essentially a centralized system, CIS was seen as "distributed" in that it could be packaged and sent out on diskettes for local use. In mid-1995, PSMFC staff began to explore options for networked data distribution, both through a dial-up bulletin-board system and the World Wide Web—"I don't like to put all my eggs in one basket," said one data manager at the time. Another difference was CIS' emphasis on quantitative detail: this design choice was partly a response to states' disappointment with NED's broad, qualitative information.

As a successor to both of these systems, StreamNet supplies both tabular and graphic information through Web-based query tools; its long-term goal is to link together decentralized information maintained by agencies throughout the region, but for now all of its information resides on a single server at *www.streamnet.org*.

5. Information sharing characteristics

NED and CIS can be characterized by the kinds of information sharing they supported, the size of each information infrastructure, and the quality of information sharing achieved.

a. Forms of information sharing

Information sharing in NED and CIS took a few, mostly simple forms. Many NED users obtained hard-copy tables and maps, or digital database or GIS extracts, from the state coordinators, Bonneville, or the Council. Many CIS users obtained digital or hard-copy database extracts from the Pacific States Marine Fisheries Commission or members of the CIS steering committee. NED and CIS were also distributed on diskettes as standalone microcomputer database applications: NED consisted of four distinct state Rivers Information Systems, a hydroelectric site database, and the Council's Protected Areas database; CIS came as a single, large information package, the "CIS Distributed System."

Many users of NED and CIS were within state and tribal fish and wildlife agencies; others were in state energy offices and water resource departments, county planning offices, and federal agencies such as the Forest Service, the Bureau of Land Management, the National Marine Fisheries Service, and the Fish and Wildlife Service. Additional users included academics, private consultants, non-profit groups, and utilities.

Generally, few of these users demanded more advanced forms of sharing, such as client-server architectures or interoperability with existing software. However, anticipating a more highly networked future, StreamNet represents a major effort at providing advanced Internet access to the region's resource information.

State	*River reaches*	*Stream miles*	*Scale*	*Source*
Montana	4,000	?	1:250k	(Montana DFWP, 1993?)
	8,000	?	1:100k	(Montana DFWP, 1994; NRIS, 1996)
Idaho	4,500	28,000	1:250k	(IDFG, 1994)
	?	103,000	1:100k	
Washington	1,500	?	1:250k	(Knudsen *et al.*, 1992)
	26,000	31,000	1:100k	
Oregon	4,000	?	1:250k	(Forsberg, 1992)
	14,000	45,000	1:100k	

Table 8-2. Sizes of state River Reach files

b. Size of the infrastructure

Another measure of these data sharing efforts was their size and traffic volume. Both data systems were large enough to warrant a detailed data-management policy and standards, yet small enough for fairly simple methods of data collection and distribution. CIS had 80,000 records, describing about 36,000 reaches. Bonneville's summary version of NED described about 34,000 reaches in all, representing 135,000 miles of streams; the individual state Rivers Information Systems varied widely by state and by scale (Table 8-2).

The traffic volume of NED and CIS can be inferred from the number of data requests reported by state and tribal coordinators. In 1995, the last year of its existence *per se*, each of NED's four state coordinators received about one request per day, according to monthly and quarterly progress reports. The Northwest Power Planning Council reported about half that figure, according to Bonneville records; thus NED's overall traffic amounted to about 1600 requests in 1995. CIS received about 300 data requests in the same year, roughly the volume of one of NED's state partners. (Unlike the other two cases, these requests are actual telephone calls logged by a data coordinator, not just hits to a server.)

One last measure: According to Bonneville's contracting officer, the yearly budget for NED was $400-$440 thousand, not including $100-$150 thousand per year to develop the River Reach File system; Bonneville's yearly budget for CIS averaged about $500 thousand over the 1988-1995 period. The combined StreamNet began with a 1996 budget of about $1.7 million.

c. Quality of the shared information

From the user's perspective, NED and CIS were only as good the information they provided: *(i)* their precision and accuracy in relation to user needs, (ii) their update frequency relative to master data sources (concurrency) and to the user's time requirements (timeliness); and *(iii)* their ease of use or encapsulation.

i. Precision, accuracy, and scale

NED's lack of quantitative precision and especially its geographic scale were its chief criticisms in its latter years. NED's initial choice of geographic scale (1:250,000) and its "soft" data (based on the judgments of local resource specialists) were fully adequate for its initial policy objectives, Protected Areas and regional hydropower planning. Designers of the Northwest Rivers Study and NED acknowledged the limitations of these choices from the outset—the 1985 GIS Task Force warned against misuse, and recommended a 1:100,000-scale update within 5 to 10 years (BPA, 1985)—although one NED coordinator saw the scale question quite differently:

> No research (to my knowledge) shows that [1:100:000 or even 1:250,000 scale data omit ecologically significant streams]. At times, I'm afraid, that excuse is used to inhibit action and change to benefit the environment. We will never know all that there is to know about our lands and waters—this should not, in my view, be used as an excuse for inaction. (...) Indeed, many of the region's old growth forests have been chopped down without the use of ANY data at all!

NED's goal was not to supply full local detail, but to provide a consistent, comprehensive view of the region's natural resources (which had never been done before). Thus, Bonneville and its state, tribal, and federal partners spent a lot of time jointly drawing up a set of assessment guidelines. As a result, NED's comprehensive "presence/absence" data on anadromous fisheries continued to prove valuable even a decade later to the National Marine Fisheries Service in evaluating the endangered status of salmon stocks in coastal Washington and Oregon.

However, as the region shifted its attention away from regional energy policy in the early 1990s, many state and tribal agencies embarked on GIS-assisted projects aimed at site-level resource management, and found NED's scale and precision inadequate. A good example of this was Washington's Surface Water Identification System, a separate set of stream codes at a scale of 1:24,000 (Vanzwol, 1995). In anticipation of these uses, Bonneville invested a steady stream of funds, starting in 1986, into upgrading the River Reach files to the 1:100,000 scale; but conflating data across geographic scales proved harder in practice than on paper. As one member of the conflation team wrote in 1993, "the result was a project that took three times longer, and cost four times more than originally estimated" (Fisher, 1993). As a result, the 1:100,000-scale products weren't complete until nearly 10 years later—by which time several states had embarked on building their own 1:100,000 or 1:24,000 scale data, with little concern for consistency beyond state boundaries. As a result, according to a tribal fisheries coordinator, several tribes in Western Washington refused to use even 1:100,000 data, claiming that it omitted 40% of the streams. This was a matter of some controversy: according to one federal fisheries manager, the 1:24,000 dataset was so dense as to be nearly unusable, featuring some streams no larger than roadside ditches. To further complicate the problem, part of the River Reach File identifier was based on a linear "River Mile" referencing system—these numbers became meaningless when transferred to a different geographic scale.

CIS' focus on tabular, quantitative data reflected some of these criticisms of NED. First, CIS was based on the same River Reach identifiers as NED, and thus corresponded to the same geographic scale, 1:250,000; but its purely tabular structure allowed easier insertion of data for smaller stream reaches. Thus, it circumvented many of the difficulties of spatial accuracy and topology implied by a GIS. Tying these newer data back to a consistent geographic framework was not a high priority for CIS' managers: they felt that GIS was beyond the reach of most of their audience, and rarely used for analytical purposes anyway. A second way in which CIS reflected NED's criticisms was its focus on collecting and providing quantitative data, instead of value-laden analyses or assessments. However, one NED participant offered a different perspective:

> There is always more quantitative data on anadromous fish because that's always where the money has been. BUT, don't be deceived. Much of the 'science' on anadromous fish is soft at best—why else the 15 year argument over how to bring them back?

The new StreamNet project adopted the 1:100,000 River Reach system as its primary geographic reference. A 1995 StreamNet proposal indicated that this referencing system would be enhanced in the future by "appropriate links to other scales." It also chose to focus its data collection efforts in the first few years, to provide greater detail on fish distributions and habitat before branching out to broad ecosystem data.

ii. Timeliness and concurrency

Fish and wildlife phenomena in the Columbia River follow a seasonal frequency, as the various fish runs make their way through the river system to and from the ocean. CIS was updated once or twice a year, as figures came in from state and tribal coordinators and other sources at the end of each fishing season. This was the frequency needed for CIS' primary use: assessing the progress of the Council's Fish and Wildlife Program. For ecological assessment, long-term trends also needed to be distinguished from short-term fluctuations, and current conditions compared with historical figures. Both CIS and NED supported this category of use, mostly at the policy level: characterizing species habitats or development suitability for the long term, with reassessments every few years. Some of NED's data categories were updated as funds and cooperative relationships allowed, and served these purposes well.

Bonneville did continue funding data-management activities in the states and tribes; but as region-wide concerns faded from view, states and tribes turned their attention to local priorities, as noted above, and left many of NED's data categories unchanged after the 1984-86 Rivers Studies. Thus, much of NED gradually became obsolete for most ecological assessments: in the eyes of many potential users, this was NED's third major limitation (after its scale and qualitative nature). The lack of updates was closely related to the first two limitations: states saw little use for NED in their site-specific work, so they saw little value in updating it; and Bonneville never directly required them to do so. One of StreamNet's initial objectives was to update several key NED categories, such as anadromous fish distribution.

iii. Usability and encapsulation

The most common mode of use for both NED and CIS was a telephone call to one of the state, tribal, or regional coordinators, who would look up the information to answer a question, or prepare a custom list or map—so measures of usability or encapsulation were not all that meaningful. NED and CIS were also distributed as stand-alone microcomputer database applications, directly usable by most recipients upon installation on a local microcomputer. More advanced uses, such as linking to local GIS systems, were not supported, though advanced users with geographic River Reach Files could fairly easily tie data records to spatial features. In late 1995, the US Geological Survey's Oregon District office began to facilitate this use by putting 1:100,000-scale River Reach files in Arc/Info format on its Web site. Around that time, the new StreamNet system also went online, with the primary goal of providing a variety of forms of Internet access to geographic resource data, in particular through Web-based mapping and querying tools.

d. Quality of the information infrastructure

In contrast with the *users* of an information sharing infrastructure, its *maintainers* generally have quite different ways to assess its quality. Their focus would be on aspects of the infrastructure itself: *(i)* its capacity for two-way information sharing (reciprocity), *(ii)* the flexibility to grow the infrastructure without overloading data management procedures (scalability), and *(iii)* its non-intrusiveness, the degree to which information sharing can adapt to participants' various procedures and standards.

i. Reciprocity

NED's and CIS' design depended on several data providers feeding back information to a central "hub" (Bonneville or the Pacific States Marine Fisheries Commission), for use there and redistribution to users throughout the region. In practice, because of its infrequent updates, NED saw few opportunities for data transfer between Bonneville and state coordinators, and so its reciprocity was only potential. In contrast, CIS was updated about twice a year, thanks to a small team at the hub which collected and integrated data from steering committee members, mostly by manual means. In 1994, the CIS team began to streamline the data integration process by means of a simple dial-in facility and standard formats for data interchange. It also began to distribute CIS data on diskettes (the "Distributed System"), and in late 1995 devised a simple Web-based query form, a precursor to the query tool now on the StreamNet Web site. Many of CIS' users were its own data providers; therefore some described CIS as a "crossroads" allowing fisheries specialists to share information crucial to their work. Around that time also, the USGS' Oregon District office put its River Reach File Clearinghouse on the Web, with provisions for coordinating updates and corrections suggested by Internet users—thus further blurring the distinction between "user" and "provider."

ii. Scalability

NED was a mostly fixed set of information and participants. But by delegating nearly all data responsibilities to the state coordinators, NED's design lessened the data-management burden at the Bonneville "hub," thus enabling some scalable growth to occur. The bottleneck in its design was the cross-reference tables that linked River Reach File numbers to the river codes used by the four states. It took one person about a year to build these cross-reference tables—and they soon became obsolete as state identifiers changed.

CIS was a rapidly growing resource, as its data managers added data from each fishing season and related them to existing records to form time series. Yet because CIS was centrally managed, its growth required a proportional increase in the data management effort at the PSMFC "hub." By 1995, CIS had grown to about 80,000 records, and its managers were devising "filter programs" to extract data from state databases, encouraging electronic data submissions in a common exchange format, and even standardizing data collection methods. These developments, still in progress when CIS became part of StreamNet, were no doubt continued as part of that project.

iii. Non-intrusiveness

In principle at least, participants in both NED and CIS were expected to follow certain conventions in managing their data, but were not required to restructure it all. As mentioned earlier, in keeping with the Northwest Power Act, NED was a loosely-integrated regional database that complemented, but did not replace, existing state activities. In particular, states could continue using the river coding schemes to which they were accustomed; cross-referencing with River Reach File numbers made it possible to assemble a consistent regional view when needed. However, once hydroelectric development became less of a crisis issue, NED's very hands-off policy made regional consistency difficult to maintain: update schedules, data formats, definitions, and stream identifiers diverged as each state pursued its own priorities with little or no common direction. Bonneville staff would reconcile the data manually in order to produce regional maps and summaries used mostly by federal agencies such as the US Forest Service and the National Marine Fisheries Service, who continued to find NED "a gold mine," in the words of a NED coordinator, for their assessments of spotted-owl and salmon habitats. However, the states and tribes, in many ways the backbone of NED's design, didn't value these course-scale, qualitative, outdated data nearly as highly; and NED gradually became simply a program whereby Bonneville funded the data management activities of fish and wildlife agencies. As one long-time observer put it in 1996, "Bonneville could take a greater role in defining data standards; but right now it just writes checks."

CIS began in an even more non-intrusive fashion than NED: it would accept anadromous fisheries data in any form, and counted on a central data management staff to reconcile it all. This policy was devised in the early days of CIS, because several key agencies (most notably the Fish Passage Center and Idaho's Department of Fish and Game) were reluctant to share information or to change how they kept their data. As a result, when CIS was transferred to the Pacific States Marine Fisheries Commission in 1993, it consisted of an assemblage of completely disparate data. Over time, CIS staff tied these data to stream reaches and began to define standards for data definitions, identifiers, and collection methods, thus demanding some changes on the part of data suppliers.

A major challenge for the new StreamNet is to define flexible standards that will let data contributors manage their data in ways useful to them, while still enabling summaries and comparisons across the region. One change that will facilitate this process: StreamNet is specifically targeted at meeting regional data needs in support of the Council's Fish and Wildlife Program. Compared with NED, therefore, it has less of a need to appease the four states: indeed, its steering committee features federal and tribal agencies on an equal footing with the four states.

6. Infrastructure impacts

The final measure of an infrastructure's quality is a pragmatic one: what has it accomplished? NED and CIS provide interesting examples of impacts, from enabling consistent region-wide strategies; to enhancing public participation and inter-state collaboration, to enhancing state and regional data management.

a. Consistent, region-wide strategies

NED was built specifically for a region-wide assessment of environmental quality and hydroelectric capacity. Thanks to the Rivers Study, the Council was able to persuade the region's public and private stakeholders to set aside 44,000 miles of streams (that is, 15% to 20% of the total) as "Protected Areas," and to redirect federal hydroelectric development to the remaining 200,000 stream miles. To some of those involved, this achievement was unthinkable before the states and tribes pooled their information consistently and systematically through NED. NED data and the Protected Areas decision also led to the first-ever estimate of future developable hydroelectric power resources, or "hydro supply curves," in the 1986 Power Plan (BPA, 1993). In speaking with those involved, it seemed that the promise of the Protected Areas amendment, and its aftermath, provided much of NED's early momentum and support.

After 1992, the National Marine Fisheries Service (NMFS) used both CIS and NED in applying the Endangered Species Act to anadromous fish in the region. According to one NMFS analyst, this Act requires assessing stream habitat and fish species by "Evolutionarily Significant Unit" (ESU), that is, a land unit supporting a genetically distinct and significant population. Compared with studying either individual salmon runs or the Pacific salmon genus as a whole, analysis at the intermediate ESU scale required extensive data processing and regional comparison, for which CIS and NED data proved valuable in determining the boundaries and health of species. Even though it was several years out of date, NED played a prominent role in assessing critical salmon runs in coastal Oregon and Washington, outside the Columbia River basin (CIS' geographic area).

Of course, regional summaries and joint projects are nothing new; but CIS and NED were infrastructures that allowed for more than a one-time effort and served multiple purposes. The Council's Northwest Power Plan and Fish and Wildlife Program have five-year update cycles; Endangered Species listings are reviewed periodically; all of these studies will be easier to repeat thanks to the regional consistency afforded by StreamNet's common reference system and its shared assessment guidelines. These regional accomplishments are good examples of "superordinate goals" or "killer applications" that often foster information sharing. NED's experience also illustrates what can happen in the absence of such a shared goal: although it was designed as a general-purpose system, and many agreed that it could be useful for other purposes, its regional focus became diffuse soon after energy development slowed down.

b. Enhanced public access to information

According to some participants, a key benefit of these infrastructures was better public access to information. In particular, simply having region-wide information organized and accessible in one place helped state and federal groups comply with public-review requirements. Open access to information was a high priority for the Northwest Power Planning Council in particular; but given public disclosure laws, many (especially federal) agencies saw it as a requirement; and others considered it a wise precaution. As one tribal fisheries coordinator put it, "Information—if you try to keep it behind closed doors, it gets pried out in court!" Increasingly widespread Internet access in the 1990s brought greatly enhanced opportunities to release information to the public.

For the most part, pricing and access restrictions were not a major issue in releasing information to the public: most of the region's government agencies had grown to consider data and products they built to be public goods, and freely gave them to anyone who asked. Montana and Oregon put their Rivers Information Systems verbatim on the Web; the US Geological Survey did the same with the entire set of River Reach Files. There were exceptions, to be sure: some agencies were reluctant to post data on the Web that cost a few dollars in paper form; some tribes were unwilling to release their seasonal harvest figures; and some agencies took a "fair market value" view of data distribution. Washington's Department of Natural Resources (DNR), in particular, considered its extensive geographic data collections to be exempt from public disclosure laws, and distributed its data on a for-profit basis. This policy was widely criticized; but DNR saw it as essential to its stewardship of state trust funds, and set up elaborate written procedures for determining fair market value and allocating funds from data sales. In the early 1990s, however, Washington's Commissioner of Public Lands called on DNR to "accept our trust mandate as broader than 'maximizing income' and recognize the value of ecosystems to trust beneficiaries." DNR's GIS staff began letting more data recipients "pay" for data in a variety of ways, such as helping to enhance a regional data layer by adding local detail, or paying only token amounts—in short, anything that might benefit DNR and seal a relationship. This new data policy, emphasizing broad inter-agency partnerships, came about in part through the leadership of DNR's GIS manager, who transposed this partnership model to the national level as chair of the National Research Council's Mapping Science Committee (National Research Council, 1994). In short, this "fair market value" policy stood out from the "free and open access" policy of most of the region's agencies, and generated controversy because of DNR's prominence in the region's GIS scene. But with the rise of "ecosystem thinking" in the region, fostered by projects such as NED and CIS, and by issues such as endangered species, even giant DNR moved closer to a policy of free and open data sharing.

c. Improved inter-state collaboration

These sharing infrastructures primarily benefited regional agencies: the inter-state Northwest Power Planning Council, the federal National Marine Fisheries Service, and others. To a lesser degree, they also enabled groups of people in the states and tribes to collaborate and to coordinate their work. At the policy level, the Northwest Rivers Study provided a mechanism for the states to wield a lot of power by working together. At the management level, CIS allowed upstream agencies in Idaho's Snake River basin to track salmon populations on the lower Columbia on occasion as an input to local regulatory decisions. Interstate collaboration may prove critical to the long-term success of the new StreamNet, to ensure ongoing support and input from the region's state and tribal agencies.

d. Enhanced data management: CIS and other NED offshoots

According to some NED participants, its primary impact was that it allowed people to concentrate on data management, thus enabling many other data-related projects. One data manager in Idaho stressed the learning opportunities provided by NED despite its infrequent update schedule in the 1990s:

> NED has indeed floundered in terms of its output. However, (...) what appears to be stagnation has really been a learning process for GIS analysts, biologists, and database programmers (like myself). (...) A gestation period was necessary, and NED has been lucky enough to have one.

For instance, most of the state agencies that participated in NED adopted its River Reach data structure to build datasets of their own. In fact, CIS itself, with its River Reach-based data architecture, was a spin-off of NED. The new StreamNet is yet another stage in the learning process, deriving from both NED and CIS and adding significant networked functionality.

7. Challenges and lessons

Both NED and CIS faced several important challenges, and learned important lessons, in the areas of decentralized coordination, data connectivity, and organizational teamwork.

a. Decentralized vs. centralized coordination

A key challenge to any partnership is maintaining a coordinated overall focus despite differing agendas and styles among participants. NED's decentralized coordination, mandated by the Northwest Power Act, worked well prior to the Council's Protected Areas amendment, which was a regional concern of direct interest to each state and tribe. With the Protected Areas milestone behind them, Bonneville and the Northwest Power Planning Council never did follow up with another compelling shared objective for NED. This failing was related to funding: NED's funds came in large part through the Council's Power Planning budget. As the region's focus on hydropower planning dwindled in the 1990s, so did Bonneville's NED funding to states and tribes, and thus its ability to guide their data-management activities in a common regional direction. "It all comes down to money," says an Idaho GIS manager. "Bonneville paid people to cooperate. No money, no conductor." Yet there were other factors: for instance, without the severe delays in upgrading the River Reach File standard to larger scales, it may well have been easier to maintain greater data consistency among a dynamic, increasingly sophisticated set of participants.

CIS had a narrower scope than NED, closely related to the salmon crisis, an issue of great concern in the region. Its support and participation may have derived from its focus on an unmistakably inter-state issue, which due to the migratory patterns of anadromous fish, no one tribe or state could resolve on its own. This interdependence among states was much less strong for NED's broader wildlife and fisheries focus. Here again, funding played an important role: CIS was funded out of the Council's Fish and Wildlife budget. Unlike the Protected Areas and regional power planning, CIS' *raison d'être* has remained prominent and will persist for a long time to come: in the words of one consultant, "Restoring endangered fish species will make the spotted-owl controversy look like child's play."

b. Providing data connectivity among data suppliers

Whether its participants knew it or not, NED was clearly ahead of its time in the mid-1980s, seeking to reconcile distributed data sources "as is," without requiring them to conform to a single homogeneous standard. This approach has been the subject of extensive research since then, leading to modern heterogeneous multi-databases, which link component data sources via a single networked interface. NED's cross-reference tables, linking each state's stream identifiers with the River Reach numbers, were a simple, yet powerful form of metadata—a term which only became common parlance in database management in the mid-1990s. Thus, even without networked connectivity, NED illustrated well the potential of non-intrusive data sharing. The difficulties that it experienced - maintaining meaningful connectivity between independently changing formats, definitions, and geographic scales—remain among the key challenges of distributed database systems.

CIS was in part a response to NED's difficulties. To facilitate comparisons, it contained quantitative data and specified how the data were measured, and established some conventions on field names and category thresholds in the data submitted from states and tribes. Furthermore, geographic scale issues were less prominent in CIS because its simple tabular structure could more easily accommodate additional river-reach records. However, with no geographic component, its use for graphic display or spatial analysis was limited to those with their own copies of the River Reach geographic database.

In merging NED and CIS into StreamNet, Bonneville, the Council, and their partner agencies sought to reap the benefits of both: NED's decentralized structure, its links to geographic features, and its support for regional policy and planning; and CIS' tight data integration and its capacity for scientific and management analysis.

c. Teamwork and coordination

Partnerships such as NED or CIS depend on a sense of teamwork among participants — a work relationship in which credit for accomplishments goes to the team, and not to any one agency. Yet, as one IRICC participant learned, most corporate incentives tend to favor leadership, to the detriment of teamwork: "some agencies have begun to submit solutions, rather than issues, to the team: the further we get from the President's Forest Plan, the more the teamwork falls apart." One participant in both NED and CIS reflected that "people who mistrust others, fear change, or don't want to share power—these are the people who undermine the collaborative process." Bonneville sought to achieve teamwork and shared ownership by giving states a lot of leeway in building NED; but as Bonneville's "generous conductor" role diminished, the states on their own had less of a regional focus once their attention turned from NED's ecosystem-wide concerns to specific anadromous fisheries. Also, as Bonneville's funding diminished, face-to-face meetings among NED's coordinators dwindled to about once a year, with quarterly written status reports. In contrast, CIS was able to maintain its "people network" through bimonthly meetings of its steering committee and frequent communications between CIS staff and other agencies (Allen *et al.*, 1994). StreamNet has maintained this bimonthly schedule for its steering committee.

Besides coordination *within* NED and CIS, two key questions facing these infrastructures was whether and how to coordinate them with each other, and with related regional information sharing programs in the region (IRICC, the EastSide Assessment, and others). The new StreamNet has begun to tackle these questions by merging CIS' and NED's data and their approaches, and by including IRICC habitat data.

8. Conclusions

The Northwest Environmental Database (NED) and the Columbia River Coordinated Information System (CIS) provide rich illustrations of both the feasibility and the challenges of decentralized geographic and environmental information. Whereas most such systems in existence today have only begun the process of putting designs into practice, NED in particular was in place for over a decade. Their experience confirms the tangible benefits of information sharing infrastructures: declaring Protected Areas in a regionally consistent, widely acceptable manner was unthinkable prior to NED, and both NED and CIS helped to replace fruitless debates over endangered species with regionally consistent, widely accepted evaluations of species status and habitat.

NED and CIS also highlight the importance of a flexible, adaptive approach to designing, implementing, and expanding information sharing infrastructures. NED's information architecture, advanced though it was for its time, was difficult to upgrade to larger geographic scales and to increased use of advanced information systems among its state counterparts. After NED's experience, CIS coped with its complex technological, scientific, and organizational context by centralizing data management responsibilities and by avoiding geographic information or scientific assessments. It could afford these choices because it was much more narrow in its focus, and it drew its funding and support from fish and wildlife protection budgets rather than power planning budgets. Now, with substantial new funding, StreamNet is trying to build on the experience of both CIS and NED.

NED's loss of regional consistency since the Protected Areas amendment illustrates the importance and difficulty of maintaining a compelling shared goal and an effective shared standard: by several accounts, this amendment marked the end of NED's enthusiastic participation and the beginning of its divergence into state-specific systems, accelerated by the lengthy delay in creating a new River Reach standard. CIS' institutional support, despite its technological simplicity and high manpower requirements, suggests that a compelling, clearly shared goal can motivate quite large investments of money and time in shared information. (In fact, the comparison is a difficult one given that CIS as a data system was only around for about three years before being merged into StreamNet.) The contrast between CIS' and NED's pattern of growth also suggests that building a community of collaborators, as CIS did through its Steering Committee, may be important to institutional support.

On a technological note, NED confirms the feasibility of "minimal standards" in geographic information, by achieving high-quality, non-intrusive information sharing using cross-references to the EPA's River Reach Files. In its "gestation period," NED also had an unseen impact as a "seed" for new infrastructures, as illustrated by CIS and other data systems which learned from the NED experience.

In summary, both NED and CIS made impressive strides in providing regionally consistent information on the natural resources of the Columbia River Basin, and in helping to defuse a very antagonistic policy context. Their data management choices were advanced for their time, and the difficulties they encountered are still those faced by much more modern and sophisticated systems today. Moreover, their ten-year history provides a glimpse into the future of today's advanced, networked systems, and offers lessons that others ought to follow closely.

Chapter 7: Case Study Synthesis

1. Salient findings from the three cases

Chapters 4, 5, and 6 presented three cases of information infrastructures being built to facilitate collaboration among government agencies: the Great Lakes Information Network (GLIN), the Gulf of Maine Environmental Data and Information Management System (EDIMS), and the interrelated Northwest Environmental Database (NED), Coordinated Information System (CIS), and StreamNet. All three chapters used the same set of criteria, grouped as follows:

- their history and context;
- choices made in building the infrastructure;
- characteristics of the information and infrastructure;
- impacts of the infrastructure; and
- challenges and lessons encountered.

The table on the next page summarizes the findings along these criteria in each case. The table highlights wide variations in duration (from GLIN's 4 years to the Pacific Northwest's 13 years), size (order-of-magnitude variations in data holdings, traffic volumes, and budgets), and institutional support (from giant Bonneville Power Administration to fledgling Gulf of Maine Council). Nonetheless, several points of comparison can be made along each of these criteria.

a. History and context

From a policy viewpoint, all three infrastructures were intended to facilitate joint protection of a water body shared by several states or provinces; but this goal varied in clarity and urgency between the cases. In the Great Lakes case, several related agreements and policies dating back several decades had articulated and instilled norms of Great Lakes stewardship and community among many people and organizations in the region. Furthermore, the US Environmental Protection Agency adopted an emphasis on ecosystem management and partnerships in the early 1990s. In addition to these norms, however, two additional elements were key to getting GLIN

	Great Lakes: GLIN	**Gulf of Maine: EDIMS**	**Northwest: NED / CIS/ StreamNet**
Lead organizations	Great Lakes Commission (GLC), Environmental Protection Agency's (EPA) Great Lakes Nat'l Pgm. Office (GLNPO), state & province agencies. Environment Canada, Consortium for Earth Science Information Network (CIESIN), others on Advisory Board.	Representatives from state & province agencies, Environment Canada, and Univ. of NH.	Bonneville Power Admin., NW Power Plng. Council, Pacific States Marine Fisheries Commission, state & tribal fish & wildlife agencies. US Forest Service, US Geological Survey, Nat'l Marine Fisheries Svce., Indian tribal agencies, and others are also on StreamNet steering committee.
History	Began in 1993 as a pilot project; grew in size and usage under several grants. Slowly moved from gopher to WWW (1994-5) and (1995) began to build in search tools, online mapping, Gov't Info. Locator Svce. (GILS) atandards.	Began in 1991 with the new Gulf of Maine Council. 1993: Built a regional data directory, with Internet access. After unfunded hiatus (1994-5), lost much of its former support.	Began in 1984 with NW River Study. 1985 GIS Task Force led to decentralized geographic design. Protected Areas in 1988. Fragmentation as funding decreased in 1990s and new River Reach files delayed until 1995. In 1996: new funds, StreamNet.
Infrastructure context: Institutional	1968 Great Lakes Basin Compact. EPA's move to ecosystem planning and partnerships. USGS, NOAA minutes from GLC in Ann Arbor, MI.	1990 Council on the Marine Environment. Oceanographic centers & associations. Atl. Coastal Zone Info. Steering Committee.	NW Power Act (1980), President's 1993 Forest Plan (US Forest Svce. And Bureau of Land Mgmt.), state Geographic Info. Councils, informal GIS communities
Technological	CICnet minutes from GLC. Widespread Internet access. EPA data projects. Corps of Engineers data ctr. Env. Canada's GLIMR. CIESIN research. Michigan's IMAGIN.	Low Internet usage, but growing rapidly in some parts. State and province GIS data centers. Oceanography research. ACZISC.	Widely varying experience. At first, little use of GIS and networks; rapid increase in the 1990s. NW Land Info. System Network. US Forest Svce. & BLM data dev't (IRICC, EastSide Assessment, PACFISH).
Infrastructure choices: Institutional	Advisory Board represents state, regional, federal gov't agencies, nonprofits, private firms, industry associations, universities, foundations.	Housed at Univ. of NH. Guidance by a Data & Information Management Committee representing state/province agencies.	*NED:* interdependence across states and sectors. *CIS:* neutrality through PSMFC. *StreamNet:* both at once, Steering Committee represents state, tribal, federal partners.
Technological	Early thrust to get people online, even if by dialup to *great-lakes.net*. Later, emphasis on indep. servers, multi-site searches, mapping, GILS.	Database directory intended to lead to decentralized data (not there yet). No choice of network protocol for several years.	*NED:* Decentralized, qualitative, geographic emphasis, coordinated thru River Reach IDs. *CIS:* centralized, qualitative, tabular. *StreamNet:* both, plus data services.
Forms of data shared	Http access to fact-sheets & agency pages. E-mail lists. Some partners have regional hydro. and economic datasets.	Http access to ocean datasets; regional database directory; documents. E-mail lists.	NED/CIS: PC database applications, and paper tables or maps. StreamNet: http interactive access to tables and maps.
Size of infrastructure	*www.great-lakes.net* saw 67,000 hits (442 Mb) in Apr. 1997. Some GLIN partners have large sets of geo. info. (30-300 Gb). Budget: $150-200,000/yr. (RAPIDS: $1 million/yr.)	*gulfofmaine.unh.edu* saw 1,500 hits (21 Mb) in April 1997. Budget: $50-$100,000/yr.	In 1995, NED/CIS reported about 180 total requests per month. Launched in 1996, *www.streamnet.org* saw 20,000 hits (188 Mb) in April 1997. Budget: $1 million/yr. (including data collection)
Information quality: Precision & accuracy	Great Lakes Commission only works with "holders of high-quality data." Info. reviewed frequently by Great Lakes Commission. Time-stamp and contact on each page.	Datasets have free-text documentation.	NED: 1:250k qualitative data grew inadequate for many users. CIS: tabular data sidestepped scale issue; quantitative data to suit some uses. StreamNet: quantitative & qualitative, 1:100k, w/ links to other scales.
Timeliness & concurrency	Mostly static data (agency fact sheets); event information updated as needed by full-time GLC staff.	Some datasets updated daily, but database directory is years out of date. Other data are static.	Parts of NED fell far out of date after 1988's Protected Areas amendment. CIS: updated twice a year.
Usability & encapsulation	Work underway on MapInfo web interfaces and multi-site search.	Some search interfaces; but most data are ASCII tables.	NED/CIS: manual distribution; StreamNet building WWW map/query interfaces
Infrastructure quality: Reciprocity	Built-in to GLIN's Web design, though sites were slow to go online. Several sites providing data now.	Despite distributed design, few servers are online other than the central server at Univ. of NH.	Built-in to the design of NED, CIS, and StreamNet: the "crossroads" metaphor. StreamNet still mostly one-way.
Scalability & flexibility	Non-scalable early emphasis on connectivity and training. In 1995: move to greater autonomy by partners.	Limited: manual, centralized data management	NED: distributed, thus scalable, except for cross-ref. Tables. CIS: not scalable until it defined data standards within StreamNet.
Non-intrusiveness	Some GILS compliance; also, HTML design guidelines to aid navigation.	Few standards at all	NED's cross-reference tables provided a hands-off stance towards states. CIS and StreamNet starting to define standards.
Infrastructure impacts	Few policy impacts; a lot of informal communication and learning. Some RAPIDS facilities hosted by GLIN.	Minimal policy impacts; informal communication & learning	Protected Areas, hydroelec. supply curves, endangered species review, collaboration and public access, enhanced data mgmt.
Challenges and lessons	Difficulty of maintaining a large Web of information resources. Need to adapt to other data sharing styles.	Need for large technological expertise, organizational support, funds, and a clear shared goal.	Importance of a clear shared goal, evolving standards, and long-term learning path. Difficulty of multi-scale basemaps.

started: a few people at the Great Lakes Commission and CICnet who articulated the relationship of these norms to shared information; and the deep pockets of the Ameritech Foundation.

In contrast, Gulf of Maine regional policies consisted only of a 1989 Agreement between three governors and two premiers; joint stewardship of the Gulf of Maine was not yet a strong or clearly articulated norm among most of the region's agencies or inhabitants. It's not too surprising, then, that participants' commitments to Gulf-wide concerns were smaller, including their ability to articulate shared information goals or to obtain and keep funds for information-related activities.

In the Pacific Northwest, the 1980 Northwest Power Act provided a clear set of norms (state-federal cooperation, and the electric power industry's responsibility to mitigate the impacts of dams) and allocated large resources from the region's electric power sales. These norms and resources grew even more significant after the Endangered Species Act was brought to bear in 1990. The twin crises of balancing electric power against fish and wildlife and restoring threatened salmon species provided a strong impetus for information-based cooperation between states, tribes and federal agencies; but the information infrastructure choices still depended on the 1985 GIS task force to articulate them, and on a few regional coordinators to keep them on the budget year after year.

All three cases, then, illustrate the importance not only of a shared ecosystem, but of a convergence between (i) shared norms about stewardship of the ecosystem, (ii) significant fiscal resources available, and (iii) someone to articulate the relationship of those norms to shared information, and to harness available resources.

On a technological note, all three cases featured an infrastructure trying to define its standards and strategies amidst a field of other, competing choices. This theme is evident in the Great Lakes case, in which GLIN was trying to define its relationship to CIESIN, IMAGIN, and others; it also appears in the Pacific Northwest case, in which NED and CIS were funded simultaneously by the same Bonneville division, and several state systems had their own data standards and ongoing GIS projects. In the Gulf of Maine, as the Web suddenly appeared on the "radar screens" of many in the Council on the Marine Environment, it suggested many alternatives for communication and information sharing, and thus not all saw EDIMS as the Council's principal mechanism for networked information sharing. These examples suggest that multiple competing alternatives to information sharing in a region may not be unusual; and that they may or may not be harmful to the effective growth of information sharing infrastructures. In a cross-section of the three cases, proponents of GLIN, NED, and EDIMS were all concerned about alternative efforts grabbing the attention and resources (both money and information) of their participants; but actual outcomes may be mixed. The many alternatives may diminish funding and institutional support (as shown by EDIMS in the Gulf of Maine). However, long-term benefits may include learning from the alternative efforts (StreamNet shows what NED's proponents learned from CIS' focus on detailed data management), or a more focused view of the infrastructure's audience and purpose (GLIN sharpened its focus on training agencies to maintain online servers).

b. Infrastructure choices

In the three cases, the primary institutional choice was an interagency committee or board to coordinate infrastructure-building efforts. The Great Lakes cases featured the largest and most diverse board, with participants from two dozen government, industry, academia, and non-profit organizations. Coordinating committees in the Northwest case featured only government agencies; NED mixed energy specialists with fish and wildlife specialists; CIS' and StreamNet's committees narrowed the focus to fish and wildlife-related agencies (StreamNet brought in the US Geological Survey to oversee Reach File coordination). In the Gulf of Maine, the committee that oversaw EDIMS was loosely defined, with several non-participating members: participants belonged to state and province agencies and universities, and came together based largely on their personal interest in coastal and ocean data management. The level of functioning of these committees (their meeting frequency, consistency of attendance, etc.) correlated with the effectiveness of the infrastructures they oversaw; it was hard to distinguish cause from effect.

Certain technological choices were also quite similar across the three cases. All three infrastructures were aimed at linking distributed information across the region without imposing too many changes on source data sites. In the Great Lakes and Gulf of Maine cases, the goal of not intruding on data sources led to a "laissez-faire" evolution with few or no standards and a diminished emphasis on detailed, dynamic data. In contrast, the Northwest Environmental Database (NED) team was more data-centered: it invested in building state cross-reference tables, and based its infrastructure on the River Reach standard. Recognizing the difficulty of maintaining data for regional use, GLIN and NED both began by equipping agencies in their regions: GLIN staff conducted training on using and providing networked data services; in the Pacific Northwest, Bonneville purchased GIS hardware and software for state and tribal agencies, and funded several full-time staff for data collection and coordination. (The Gulf of Maine data team, with a much smaller budget, simply deferred its distributed data plan until more agencies put servers online.)

The three infrastructures show quite different choices, however, in their emphasis on geographic information: GLIN waited until its third year to begin building interactive map interfaces on the Web; whereas EDIMS hadn't yet taken the GIS plunge after some five years of existence. As for the Northwest case, NED was inherently geographic from its start in 1985, but CIS handled only tabular data to avoid the complexities of geographic data; and StreamNet featured an online map library, later joined by an interactive "map builder."

c. Infrastructure characteristics

One salient point in common between the cases: from quite different beginnings, all three infrastructures converged onto a Web-based architecture, in which a central "hub" server was intended as an intermediate "scaffolding" for building a future distributed web of data servers in the region. However, evidence suggests that moving beyond this "scaffolding" may be difficult. The Gulf of Maine EDIMS has remained at the centralized stage for much longer than expected. The Great Lakes network was almost entirely centralized for a long time, but saw other agencies' servers come online in 1995-96. As for StreamNet, it's too early to tell how easily this so-called "Aquatic Information Network" will evolve beyond its current centralized form.

The size and traffic levels of the three infrastructures are quite different, in keeping with their vastly different funding levels, and reflect the length of time they've been in service, the size and breadth of their audiences, and the usefulness of their information to their audiences.

The quality of the information distributed by these infrastructures is difficult to compare, given the quite different forms of information shared and the different uses targeted. Nonetheless, usability and encapsulation are one clear point of comparison: much of EDIMS' data was raw ASCII tables, whose use required considerable processing and interpretation. StreamNet (and its predecessors NED and CIS) presented information to users in more usable form: database queries as well as maps (and an interactive "map builder" on the Web). Some of GLIN's partners feature similar query interfaces; GLIN staff also began to experiment in late 1996 with map query tools on the Web.

The quality of the three infrastructures is a bit easier to compare. Reciprocity (two-way information flow) was part of all three designs, but carrying out that vision on the Web has proven difficult given that few sources of data could or would serve it online. All three infrastructures are scalable in that they can link to new Web sites as needed; but none of the three has achieved a scalable data management strategy. As for non-intrusiveness, all three infrastructures recognized the need for it, but GLIN and EDIMS went so far as to define no standards at all. GLIN eventually defined design guidelines and a keyword list, and began to comply with the Global Information Locator Service (GILS) standard to facilitate document search and retrieval. The three infrastructures in the Pacific Northwest were more data-centered, and adopted the River Reach files as their common standard from the outset. NED was an early implementation of non-intrusive design; as StreamNet moves out of its centralized "scaffolding" form, it may need to re-visit NED's non-intrusive approach and to adapt it to the networked environment.

d. Infrastructure impacts

Users of all three infrastructures report learning experiences and improved collaborative relationships. The Great Lakes has the most tangible evidence of this, with a great many servers going online, and large traffic volumes on GLIN's core server. In the Great Lakes, the Regional Air Pollutant Inventory Development System (RAPIDS) is claimed as a GLIN impact—but it was really a separate EPA-funded project, directed by GLIN's director, that used GLIN's main server to exchange electronic mail and to interchange software under development. Only NED in the Pacific Northwest seems to have *already* had clear, tangible impacts on environmental planning and policy: impacts in the other two cases are mostly in the future tense, or in less specific, measurable terms.

e. Challenges and lessons

One challenge encountered in all three cases is the complexity of building and maintaining networks of distributed, dynamic information. Building the "scaffolding," a temporary structure to get started, was relatively easy; but moving on to a distributed data architecture, or to dynamic data rather than fixed "brochures," or maintaining an infrastructure over time, was more difficult. Related to this is the importance of large commitments of resources, considerable technical expertise, and especially a technical and institutional growth path. Indeed, in none of the three cases was there a comprehensive blueprint that could be followed to build the mix of network technologies, data structures, and organizational support required for an effective infrastructure; GIS and Web-based software tools were subject to especially unpredictable change. Therefore, the process of improvised learning and growth was more important than any set of factors that could have been established in advance. Yet at the same time, a completely improvised, "laissez-faire" approach, with too little attention to the design of organizational, technological and policy systems, proved less than effective as well. Section 4 below explores in more detail the dilemma of balancing the planned and improvised approaches.

2. Differing views of information and sharing

One way to generalize beyond these cases towards a more general pattern of infrastructure design, growth, use, and impact, is to examine what has motivated these people to act in their various ways—that is, their perspectives on the problems to be solved by information-sharing infrastructures and apparent solutions they provide. These multiple perspectives are possible because information-sharing infrastructures are currently in a state where many choices are possible, and many definitions of the problem coexist: a state of *interpretive flexibility* (Pinch and Bijker (1984). What follows is a sketch of the salient views of information sharing, the designs they engendered, and what these different views have to offer both separately and in concert.

a. "We will serve no data before its time!": Information sharing as data management

The tongue-in-cheek quote above was from a hydrologist at the U.S. Army Corps of Engineers in Detroit, which maintains several hundred gigabytes of data on Great Lakes hydrology, bathymetry, and meteorology. While this quip was a deliberate exaggeration, many keepers of large geographic data collections, across the three cases, did hold the view that lengthy preparation and documentation of information resources was both necessary and sufficient for inter-agency data sharing: at Michigan's Resource Information System, technicians worked night and day shifts to manage some 40 Gb of vector data sets, and distribute them through a service bureau. A GIS manager at the U.S. Environmental Protection Agency (EPA) in Chicago painstakingly documented all geographic information before releasing it, and compared himself to a "sea anchor" preventing EPA staff from being swept away by a data-hungry public. The state of Washington's Department of Natural Resources (DNR) put a lot of emphasis on data management, license agreements, and pricing procedures, in keeping with their view of data as a state asset alongside timber or minerals. A chief concern in all these settings was to protect the data investment by releasing only high-quality, extensively documented data and ensuring that it was not misused. The resulting infrastructures tended to consist of quite formalized administrative procedures (price sheets, Memoranda of Understanding, and signatory partnerships). Also, maintainers of such infrastructures were slow to use data networks: to them, the Internet brought mostly security holes, and chances to "get in over one's head." For instance, Michigan's sophisticated Resource Information System (MIRIS) awaited a statewide networking plan for months before experimenting with any kind of Internet data distribution: "otherwise, we'd be alone in figuring it out," said one MIRIS administrator. (Noticeably absent from these characterizations is the Gulf of Maine EDIMS, which hasn't yet tried to tap into the sizable repositories of geographic data in the states and provinces.)

b. "Democracy in action": information sharing as public disclosure

Through the above quote, an editor for the Great Lakes Information Network (GLIN) emphasized the goal of making information freely available to the public—supporting inter-agency collaboration was not the primary goal, but an immediate and fortunate consequence of communicating with the public. In keeping with this view, a former chair of GLIN's advisory committee argued that "information is the only commodity worth anything today; withholding it results in costs to society." In the Pacific Northwest, a tribal fisheries coordinator had concluded from that region's experience that it was foolish to keep information "behind closed doors." And although the Gulf of Maine data team hadn't made public disclosure a high priority (their focus was state and province agencies), they were getting pulled in that direction by the Gulf of Maine Council's committee on public education and outreach. These groups all held high the public disclosure of data, above concerns for privacy, documentation, expense, or direct personal or agency benefit. The resulting information-sharing infrastructures were informally structured, and targeted a broad audience and simpler forms of data. They did much to turn public agencies outward towards interaction with the public. But their emphasis on public channels risked overlooking collaborative opportunities that involved more confidential data; or excluding partner agencies that didn't have the same openness towards public information and participation.

c. If we build it they will come: information sharing as a networking project

To some participants in the cases, information sharing was chiefly a matter of building networked data access tools. In the Great Lakes case, two foremost goals of the Consortium for International Earth Sciences Information Network (CIESIN) were to test its novel "Gateway" software and to grow its Socio-Economic Data Archive Center. This meant an emphasis on building prototype data structures, query tools, and data catalogs. On a smaller scale, the Gulf of Maine's Environmental Data and Information Management System (EDIMS) had a similar emphasis on putting data on a website and counting on people to fetch them: the information itself was not packaged for any particular use, and its relevance to users' tasks or needs was not yet a primary concern. In the Northwest case, NED and CIS were set up long before Internet connectivity became widespread among the region's government agencies: without the distractions of a networked environment, their designers could focus on data-centered design choices. Once these systems were in place, StreamNet's network-intensive approach proved slow to implement, and required going beyond incremental changes (such as the CIS "distributed system"). But thanks to its data-intensive predecessors, StreamNet could draw on a long technological and organizational foundation of joint data collection and management throughout the region, and quickly go beyond a networking project to serve its audience with useful, valuable data.

d. "We learn from each other": information sharing as a meeting of the minds

For many participants in these cases, the clearest payoff of the information-sharing infrastructures was fresh ideas from people with similar professional interests in distant organizations, whether through electronic mail or Web pages. Granovetter (1972) emphasizes this "strength of weak ties" in diffusing information through a social network. In the cases, informal learning was a frequently-cited benefit of all three information infrastructures. Members of Great Lakes agencies, for instance, could keep in touch with regional events, funding opportunities, and the activities of their counterparts in other states or across the Canadian border. In the Gulf of Maine, coastal-resource managers could learn about environmental conditions and different regulations in other parts of that region. In the Pacific Northwest, several participants report learning from people on other sides of that region's environmental debates: fisheries managers, electric-power planners, foresters, and so on. Some in that region also report learning about information management and GIS through NED and other inter-state data sharing efforts. Within this view, networked infrastructures exchanged primarily knowledge and expertise, rather than specific sets of data. Building professional communities is an important part of supporting inter-agency collaboration through information-sharing infrastructures. However, absent a common base of shared information (as in the Gulf of Maine case in particular), each participant may learn new things but only the most informal of joint actions is possible.

e. Getting it right: creatively solving new problems, organizational change, integrated choices

The above are some of many views of the "information sharing problem" observed in the cases: *(i)* painstaking data management, *(ii)* public disclosure and participation, *(iii)* networked experimentation, and *(iv)* professional cross-fertilization. These views yielded an equally wide array of technological solutions: *(i)* data standards and procedures, *(ii)* open-access sites, *(iii)* custom networking tools, and *(iv)* electronic forums.

All of these views are important and have something to offer to the design of shared information systems; and most of the time, people can work on these different areas in isolation from each other. For instance, rigorous data management creates a trusted information resource, one that people can count on to be available and reliable; public disclosure encourages organizations to be more open to new relationships and opportunities; networked experimentation is needed to stay abreast of a rapidly-changing set of technologies; and forming professional communities is important to keep everyone's choices mutually beneficial. However, at certain decision points in the growth of the infrastructure, too sharp a focus on any one of these may obscure choices that could support more than one goal. For instance, an exclusive focus on data management, through robust data structures, security, or extensive documentation, may overlook information structures that would serve both internal and external needs. Conversely, focusing purely on communication and networking may overlook the importance of robust and secure data management, or may even omit structured alphanumeric or geographic data in designing the infrastructures. The cases showed several examples of a strong dichotomy between the data management and communication functions: few instances featured data structures suitable for external communication, or communication media suitable for detailed structured data. Thus, even though these different views of information sharing, and the design choices they lead to, can happily coexist most of the time, there are times when someone needs to bring these views together, and consider multiple goals and multiple development paths at once. In the Great Lakes and Gulf of Maine cases, the slow move away from temporary "scaffolding" systems may have been due to the difficulty in bringing these different areas of expertise to bear.

A second limitation of these partial views is that trying to use a new, still-malleable tool to solve existing problems may miss an important potential of information sharing infrastructures. Perhaps the "problem" of inter-organizational information sharing is better described as a situation in which several problems and solutions can be creatively invented and reconciled (cf. Carlson, 1992). If so, then successful cases of information sharing infrastructures will be those where important new problems are identified and resolved through information sharing—rather than those where networked information is simply used to address existing problems. As Weick (1990) suggests, "one of the many streams found in organized anarchies is a steady stream of solutions in search of problems." As far back as 1980, researchers on telecommunications and other information technologies have suggested that "new computing systems are often applied not only to existing organizational problems but to qualitatively new organizational activities." (King and Kraemer (1980), cited in Attewell and Rule (1984)).

As a consequence, these views (*i-iv*) are also limited in that they presuppose an essentially unchanged organizational structure: this assumption may be incompatible with information sharing infrastructures. The "qualitatively new" activities supported by information sharing infrastructures will often require deep changes in the priorities, dependencies, and identity of participating agencies: this possibility tends to be overlooked as various participants set their expectations. Integrating these different views together into a single information-sharing infrastructure holds a lot of promise. The following section suggests a number of ways to leverage technology, organizations, and policy in order to move towards such an integrated view.

3. Mutual influence of technology, organizations, and policy/planning

As another way to interpret the case study findings, Chapter 2 suggested that sharing geographic information ought to be viewed as an interaction between organizations and technology (Markus and Robey, 1988). As detailed in that chapter, a *structuration* perspective highlights the "intertwined" nature of technological and behavioral choices (DeSanctis and Poole, 1994), and allows technology to be defined as a malleable element of organizational structure (Orlikowski, 1992)—that is, a set of rules and resources that enable some actions, while constraining others, and that are in turn shaped by those actions over time. Within this perspective, the technical design of an information sharing infrastructure is ineluctably tied to its implementation and use within an organizational context. This provides a fruitful model for understanding the processes of ongoing learning and growth evidenced in the case studies. Orlikowski (1992) traces a three-way cycle of mutual influence between institutional properties, technology, and the human agents who use and build the tools within the institution. The cases in Chapters 4, 5, and 6 suggest a similar cycle of influence between organizational, technological, and policy/planning structures, and the actions people perform within those structures (Figure 7-1).

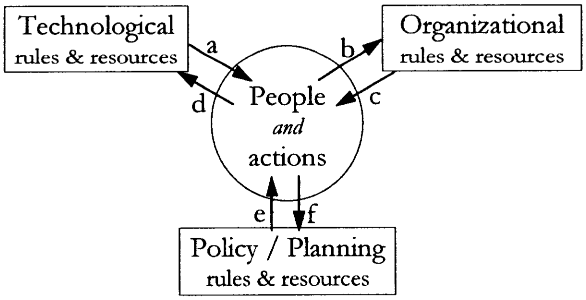


Figure 7-1. Mutual influence of technology, organization, and policy/planning on people and actions

Based on these reciprocal influence patterns, we can retire old "chicken-and-egg" dilemmas such as the role of technological *vs.* human factors in effective policy and planning. In their place, these cyclical, ongoing patterns of change provide a more accurate understanding of growth and change mechanisms. Indeed, rather than postulate direct influences between constructs like technology, organizations, or policy, this model sees all of the influences as initiated and mediated by human actors who are enabled and constrained by these constructs. This model also provides any number of "levers" for perturbing existing behavior and guiding it towards a particular target. To that end, the following paragraphs examine each of these six influences as they are illustrated in the cases, ending with a few comments on the central role of dynamic structural change.

a. Technology influences actions

Tangible impacts of technology on action (arrow **a** above) were often less than had been hoped for in many agencies reviewed in the Great Lakes and Gulf of Maine cases. Except for those who actually ran these infrastructures, most members of these agencies had kept much the same tasks, priorities, and identity they had prior to using information sharing infrastructures. However, in the Pacific Northwest, even without the grand technologies of wide-area electronic networks, NED's use of a simple regional information standard did much to influence people's actions: many who had been loath to disclose information changed their minds when they saw the political strength to be gained by helping to build a regional resource inventory based on the River Reach Files. Furthermore, NED's intuitive data categories and colorful maps proved instrumental in helping participants to articulate their priorities and (in this case) to find common ground on which to act. On the other hand, technological difficulties may also hinder changed actions: despite NED's early success in the Pacific Northwest, lingering data incompatibilities (nominal data, identifiers, etc.) and the absence of data networks for many years, this group of people had difficulty making a broader set of joint decisions in the ensuing years. In the Great Lakes, air-pollution specialists used widespread Internet access for regional pollution assessment through RAPIDS; but information differences, primarily language (English *vs.* French) and measurement units (feet *vs.* meters), did complicate other trans-boundary projects.

b. Actions influence organizations

The cases show several examples of people's actions affecting organizational structures (arrow **b** above). In the Gulf of Maine case, the first committee chair to oversee EDIMS invested time and effort to obtain federal funding, thus making information-sharing systems highly visible in the early years of the Gulf of Maine Council. In the Pacific Northwest, the efforts of NED's regional coordinator did much to encourage continued familiarity and cooperation between agencies. In the Great Lakes, GLIN's director at the Great Lakes Commission played much the same role, by keeping in touch informally with many GLIN partners and assisting many others to put information online and join the GLIN partnership. As a result, many GLIN participants felt quite strongly that exchanging networked information about programs or personnel was leading to improved coordination and responsiveness between agencies. Conversely, some felt this benefit sometimes brought with it greater outside scrutiny and a potential loss of privacy—especially when using public channels like the World Wide Web. This, too, was a (perhaps unintended) organizational effect of people's actions.

c. Organizations influence actions

The organization's influence on people's choices (arrow **c** above) takes a variety of different forms. First, the organization's size and experience clearly affect the ambition of information-sharing initiatives. It's no great surprise, for instance, that the Pacific Northwest StreamNet took on broader objectives than the Gulf of Maine EDIMS: it had twice as many participating agencies, a budget more than ten times as large, technical input from some of the region's top GIS experts, and a fifteen-year history of grappling with inter-organizational information. Organizational properties also show their influence when group "insurgencies" (Thomas, 1993) enable individual "champions" to lead others in making key choices. One such insurgency was the so-called "users' revolt" by state GIS centers in the Pacific Northwest when the first version of the upgraded River Reach files failed to meet expectations. This movement was led by two individuals, but it gained prominence because the state's GIS centers had reached a level of organizational maturity and technological expertise that enabled them to see beyond the "marching orders" of their federal funding sources. In the Great Lakes, the Great Lakes Commission's long-standing role as an inter-state compact, and CICnet's position as regional Internet provider, prompted and empowered a few people at the Commission and CICnet to launch the GLIN concept and follow it through. In the Gulf of Maine, the organizational context had a dampening effect on EDIMS development and support: due to the Gulf of Maine Council's relatively weak role in relation to the region's coastal and marine management agencies, EDIMS could not count on many agencies in the region to support its growth.

d. Actions affect technology

The cases provide many examples of people's actions affecting technology (arrow **d** above). In the Northwest example just cited above, although organizational maturity allowed two individuals to lead a "user's revolt" with important consequences on regional standards, nonetheless what actually started the change was these individuals' dissatisfaction with the product they were given, and their choice to do something about it. Similarly, in the Great Lakes case, GLIN's open-ended, decentralized design were due largely to one person's vision and enthusiasm for a regional network, combined with another person's technical abilities and shared enthusiasm. The Gulf of Maine's decentralized design owed much to the strong preferences of its hosts at the University of New Hampshire, who believed strongly in decentralized systems with few standards. Also, the slow growth of EDIMS beyond its centralized form can be attributed to the hesitancy of some members of the EDIMS committee to plunge into Internet connectivity at the expense of other networking alternatives. Although in retrospect, organizational or policy conditions might seem to have dictated these choices, nonetheless these and other individuals were key to starting and promoting (or slowing) technological change.

e. Policy and planning influences action

An agency's policy and planning environment can have important impacts (whether deliberate or not) on what its members do (arrow **e** above). Ecosystem-based policies in particular often require or encourage agency personnel to coordinate and collaborate across jurisdictional boundaries (and thus to form interagency committees and to use information-sharing technology). The Northwest Power Act is a good example of a policy that has shifted public investments strongly in favor of inter-agency relationships and shared information systems. Policy and planning goals also affect budget levels, and thus the magnitude of investments in organizational or technological change. The Northwest Power Planning Council, for instance, has a modest size (as measured by its personnel and budget), but the Northwest Power Act endows it with a lot of power to leverage Bonneville funds; its members and partners have used this position to influence the region's choices of data collection and alliance formation. The discussion of "norms" in Section 1 above discusses other examples of how the policy context has influenced people's choices in the Great Lakes and Gulf of Maine cases.

f. Actions influence policy and planning

Finally, people using new technologies within these emerging organizational structures made a few region-wide policy and planning choices (arrow **f** above)—and sometimes, the difficulty in changing technology or structures kept people from making effective policies. In the Pacific Northwest, for instance, a handful of NED coordinators using a regional information standard played a major role in drawing up the region's first system-wide regional hydroelectric and environmental protection plan. This was the ultimate goal of the information-sharing infrastructure in all three cases, although NED is the only one to have already had a clear policy impact.

g. Technological and organizational change

Besides tracing cycles of influence between structure and action, the structuration perspective also posits the potential for ongoing change in social structures. It assumes that Spackman's (1990) "organic, yet systematic" change over time is normal and ongoing, rather than surprising and temporary. Thus, as detailed in Chapter 2, information sharing infrastructures are seen not as fixed sets of interlocking components, but as a chosen *direction*, or even a *style* of evolution through an uncertain future. Change is usually a response to new technology, funding, relationships, or policies; but (whether intended or not) it can also serve to develop and maintain an organization's agility. The case studies illustrate each kind of change.

Changes in the technological or organizational context can threaten the survival of information sharing infrastructures. Recent years have brought an acceleration in the turnover of networked information technologies: barely a year after the Great Lakes Information Network got underway, for instance, it had to make major changes to its information services, switching to the World Wide Web after having put a few hundred documents on Gopher servers (recall that Gopher was the Internet technology of choice in 1993). The World Wide Web also caught the mostly telnet- and ftp-based Gulf of Maine EDIMS by surprise in 1994, just as it was being mothballed for lack of funds. When it awoke from its 18-month slumber, the "cyberspace" phenomenon was in full swing, and several committees in the Gulf of Maine Council were clamoring for a sophisticated Internet presence. The organizational front is just as unpredictable: new powerful partners or competitors may arise unexpectedly; others may go dormant; alliances may wax and wane. In the Pacific Northwest, congressional actions on regional energy supply, salmon restoration, and endangered species led to an ebb and flow of federal funds over the years; with each wave of funding came new webs of inter-agency partnerships and new attempts at data standards.

To respond to an unpredictably changing world, information-sharing infrastructures need to be thought of, and designed, as living, growing organisms rather than fixed artifacts. Management guru Tom Peters has written extensively about deliberate organizational change, "necessary disorganization" in response to a unpredictable, rapidly evolving business context (Peters, 1992). One example of this "design-for-change" was the Pacific Northwest River Reach Files: even before the first version was deployed at the 1:250,000 scale, plans were underway for a dramatically upgraded second version at a 1:100,000 a few years down the road.

In summary, then, the perspective depicted in Figure 7-1 highlights the many influences at work between technologies, organizations, and policies, and the people who live and work within these structures. Separating these different influences may seem a bit artificial in places (for in stance, it's through using networked technologies that people may be inspired to create inter-agency relationships). Nonetheless, it's important to highlight the role of (ultimately unpredictable) individual actions in building, maintaining, or changing technology, organizations, and policy. By separating out the different influences of technology, organizations, and policy, it becomes a bit easier to anticipate the effect of adjustments to these structures, either separate or in concert. Even so, none of the three cases played out a single predefined strategy, but rather evolved in response to a number of dilemmas and a shifting set of policies, organizations and technologies. More generally, then, it would seem that no strategy can have fully predictable impacts: instead strategic choices are likely to evolve with the changing context.

This uncertainty may be especially important when dealing with geographic information systems, for which the technical tools are still evolving rapidly, and which (as detailed in Chapter 2) tend to encourage relationships that cut across traditional hierarchies and boundaries and thus encourage either official or *de facto* organizational change.

Keeping these uncertainties in mind, the next two sections consider the possible role of new technologies and organizational forms.

4. Technology choices

a. From datasets to data services

Findings from the case studies suggest a number of promising directions for the design of information-sharing infrastructures. In particular, certain new technologies promise useful perturbations in the "influence cycle" described earlier. In particular, one frequent theme in the three cases was a dichotomy between data management and inter-agency communications. Each of the cases featured several Web servers, but most of these delivered only static pages and graphics; few provided a gateway into an organization's larger (especially geographic) information resources. These resources was generally kept behind database managers, GIS packages, or other software intended for interactive use by a data custodian. Accordingly, while one set of people focused on inter-agency electronic communications, another worked independently to manage the agency's information resource and distribute it on compact discs, magnetic tape, or paper maps. Quite possibly, merging these two functions would reap greater benefits from information-sharing infrastructures. Two important requirements for this are communications systems geared for queries against structured data, data management systems geared for online access, and a commitment to serve both internal and external needs in concert where possible. Newer client-server architectures may help with these requirements: the next chapter uses a simple example of a geographic data service to explore the implications of such emerging technologies.

b. From standards to metadata

Chapter 2 contrasted traditional data standards with newer "minimal" standards, in which metadata are used to help users make sense of information without altering the information itself. The cases showed many settings where this approach might be useful: independent agencies that had an interest in sharing information, but were unlikely ever to agree on a common set of definitions, formats, or coordinate systems for their data. The Great Lakes Information Management Resource provided an example of metadata for information sharing: a set of common keywords allowed a single query at any one site to retrieve relevant documents from several Web sites at once. The cases also showed an interesting early example of geographic metadata, the cross-reference tables that linked state stream identifiers in the Pacific Northwest to the regional River Reach Files. The next chapter looks more closely at the benefits and challenges of maintaining and using geographic metadata.

c. From compatibility to interoperability

If information is shared through an interaction between software systems, then getting the software programs to understand each other becomes more important than reducing data to a common format, structure, or coordinate system. This requires generic reference models of the geographic entities and operations to be shared, and program interfaces between those reference models and individual information systems—that is, interoperable data services rather than compatible datasets. Interoperability is a long-term goal of many information-sharing infrastructures such as those in the cases. However, the difficulty of defining a reference model to suit widely-varying applications, has made progress slow thus far. In addition, as illustrated by the Pacific Northwest River Reach Files, geographic information presents special challenges, such as *scale*, not raised by the alphanumeric and administrative information behind most Information Systems research. The next chapter evaluates a few small steps towards interoperability between geographic data services, and towards scale-independent geographic referencing.

5. Organizational choices

a. From autonomy to interdependence

Finally, a more "online" view of information, with data services, metadata, and interoperability, implies several organizational shifts for the agencies involved. First, it requires data managers to take responsibility for providing reliable data services to others outside. Conversely, it expects analysts and decision-makers in these organizations to depend on outside data for essential tasks. Third, learning to live with different data conventions, and to negotiate them through metadata, implies informal, yet trustworthy cooperative relationships between organizations. In fact, reaping the intended planning and policy benefits of information-sharing infrastructures may require deeper-than-expected changes in organizations and inter-organizational relationships. This may explain the difficulties GLIN proponents had in seeing tangible impacts of information sharing: most of the region's organizations had not yet undergone these deeper changes. The same explanation may have been behind EDIMS' lack of relevance to Gulf of Maine organizations. By comparison, in the last decade, many more agencies in the Pacific Northwest began (whether willingly or forcibly) to see inter-organizational coordination and information interchange as fundamental to their work, and made appropriate organizational changes: this may have helped them to reap greater benefits from their information-sharing infrastructure.

From observing the business world, Moore (1996) suggests that increasingly complex relations among organizations have made traditional market models obsolete: in their place, he proposes a "business ecosystem" ruled not by autonomy, authority, or competition, but by interdependence, persuasion, and "co-evolution" with competitor/partners in a constantly changing business context. This perspective seems at least as convincing in the world of public management as in the business world. In particular, in the complex trans-boundary contexts of the three cases, many organizations are trying to look beyond a traditional emphasis on leadership on a policy issue, and to embrace teamwork within a shifting pattern of alliances and resources. Chapter 9 explores the organizational and strategic implications of "government ecosystems" linked by information-sharing infrastructures.

b. Growing complexity over the long term

Finally, the cases suggest that building effective geographic information infrastructures may require balancing simple, tangible results early on with a much more complex permanent structure, attained through a long term path of intentional learning and growth. As mentioned earlier, none of the three cases followed a single strategic "blueprint" throughout their evolution: the technical and organizational requirements are too complex, and too context-specific for that—especially where they have attempted to handle geographic information. Chapter 9 explores this question of planned and improvised development, and the tension between incremental and more radical change.

c. Sharing costs and benefits reliably

The balancing act between simplicity (now) and complexity (later) raises another dilemma: who pays? The costs are often immediate, ongoing, and sizable; the benefits, though potentially great, are often distant and uncertain. Given this equation, the apparent "bureaucratic inertia" often encountered in geographic information sharing should come as no surprise. This theme is explored further in Chapter 9 as well.

Chapter 8: Prototype Development: a Digital Orthophoto Browser for the Boston area

1. Introduction

The preceding chapters (4-7) examined several geographic information infrastructures being built to link the activities of government agencies. This provided insights into the role of today's state of the art technologies in these agencies, and into the complexity and uncertainty surrounding the growth and implementation of such systems in such a context.

Whereas the preceding chapters focused on the complexities of inter-organizational context, this chapter adopts a very different angle on geographic information infrastructures. It reports on a year-long software development project, conducted within the broader United States National Spatial Data Infrastructure, and aimed at building tools for enhanced network dissemination of digital orthophotos. The project's outcome, an interactive browser interface to digital orthophotos[4] for the Boston area, embodies several design principles outlined earlier for loosely-coupled sharing of information. In particular, it embodies a data services model for information sharing, it provides several kinds of access to these services, and it employs metadata to facilitate networked data discovery and interoperability between software systems.

The project is a useful contribution to the discussion of geographic information infrastructures (whether national, or regional as in the case studies) in that it surfaces, first, several key tradeoffs in building the needed packages of technology. It also gauges, in a very tangible way, the impacts that may arise from assembling existing and emerging network technologies in new ways. Third, by examining the technological side of infrastructures in some depth, and by considering the role of emerging network tools, the project provides a useful counterpoint to the organizational case studies, and extrapolates their conclusions into the near future within a changing technological context.

---

[4] Digital orthophotos are aerial photographs that have been digitally scanned and orthorectified, that is, their image pixels have been repositioned along the grid of a standard geographic coordinate system to eliminate distortions from topography or camera angle. (In this case, the images are rectified to the Massachusetts State Plane coordinate system, using the North American Datum of 1983 to describe the earth's curvature). Rectification is by a mathematical trans-

The following sections summarize, first, the goals of the US National Spatial Data Infrastructure (NSDI) and how this project was to contribute to them. This is followed by a sketch of the project's specific goals and development stages; the functions provided by the final product; and possible implications of the project for the NSDI and the three case-study contexts.

2. The National Spatial Data Infrastructure and the orthophoto browser

a. National Spatial Data Infrastructure overview

The orthophoto browser is in part an outcome of several years of activity in state and federal governments in the area of geographic data sharing and standards. In late 1990, the US Office of Management and Budget established the Federal Geographic Data Committee (FGDC) with a mandate to promote and coordinate digital mapping activities throughout the country (OMB, 1990). One of the FGDC's key objectives has been to build a National Spatial Data Infrastructure: this was proposed by the Mapping Science Committee of the National Research Council (1993) and established by a presidential Executive Order (EOP, 1994). The NSDI is described as "an umbrella of policies, standards, and procedures under which organizations and technologies interact to foster more efficient use, management, and production of geospatial data" (Tosta and Domaratz, 1997). NSDI development is along three principal objectives:

The *National Geospatial Data Clearinghouse.* The Clearinghouse is a decentralized network of Internet sites that provide metadata describing available geographic information. Subsequent to the Executive Order, the FGDC adopted the Z39.50 standard and developed a metadata Content Standard (FGDC, 1994) for search and retrieval functions in the Clearinghouse. (The Executive Order also calls on federal agencies to "draw up a plan" for distributing geographic information itself to the public; but Clearinghouse efforts thus far have remained focused on metadata.)

Standards to facilitate the sharing of geospatial data. These standards are established through broad consultation with state, local, and tribal governments and the private and academic sectors. Stan-

---

formation that stretches, squeezes, twists, and bends the image based on the coordinates of known control points and on a digital elevation model that describes the altitude of each pixel.

dards to date have included the Content Standard for Geospatial Metadata (FGDC, 1994) and the Spatial Data Transfer Standard (NIST, 1992).

The *National Digital Geospatial Data Framework*. The Framework is a set of the most commonly-used layers of geographic information, including transportation, hydrology, political boundaries, and digital orthophotos. Responsibility for collecting and maintaining these data layers is shared among state and local agencies in a given region (FGDC, 1995).

Underlying these three primary thrusts, the FGDC has emphasized training programs and interagency partnerships, in recognition of the importance of decentralized approaches to maintaining detailed geographic information (Tosta, 1994). To promote participation in the first of these objectives, the Clearinghouse, in 1994 the FGDC established a Competitive Cooperative Agreements Program (CCAP), a series of small "seed money" grants to state and local partnerships to assist them in setting up Clearinghouse nodes with FGDC-compliant metadata and Z39.50 Internet access.

b. Relation of the orthophoto browser project to the NSDI

The orthophoto prototype was developed with funding from the FGDC's Competitive Cooperative Agreements Program (CCAP).[5] The project was devised to meet FGDC Clearinghouse requirements by providing standardized metadata on orthophotos via the Z39.50 protocol. In addition, it used the orthophoto service as a test-bed to experiment with alternative ways to

- provide networked access to the orthophotos and metadata by a broad class of users;
- distribute the orthophotos efficiently over a limited network bandwidth; and
- link the networked orthophoto service more tightly with client GIS software.

These additional objectives were a response to likely futures of the National Geospatial Data Clearinghouse. First, as the Clearinghouse goes beyond delivering metadata and begins to serve large datasets, efficient and manageable network access will require query mechanisms for retriev-

---

[5] Proposal abstract online at <*http://www.fgdc.gov/CCAP/CCAP95WIN.html*>. Funds awarded July 1995.

ing portions of a dataset according to geographic coordinates, resolution, and other criteria. By using image data, this project afforded a chance to begin exploring the distribution of large geographic datasets without the burden of complex or proprietary data structures. Second, to use Clearinghouse information effectively, users will often need to compare, complete, or otherwise combine it with their own information resources. This motivated the development of new forms of inter-communication between navigational Web browsers and analytical GIS viewers. Third, to foster the distributed, "organic" growth of the Clearinghouse as data providers put suitable information online, the orthophoto service was designed to be easily managed and replicated at other sites.

The project's intent was to contribute not only to the Clearinghouse effort itself, but also to inform the design and development of the NSDI's networked standards and of its Framework distribution mechanisms. As such, alongside detailed metadata, the orthophoto project provided digital orthophoto data, reusable, configurable tools for networked access to orthophotos, and insights into future NSDI development strategies in a networked context.

3. Design goals and development stages

a. Background and design goals

Digital orthophotos are useful for a wide variety of purposes, given that they combine the visual richness of a photograph with the geometric qualities of a map. This makes them useful both as backdrops to traditional line maps (zoning, transportation, etc.) in geographic information systems (GIS), and as clearly intelligible representations of geographic space for a broad audience.

However, orthophotos are usually large enough (tens to hundreds of Megabytes apiece) to exclude all but specialist, dedicated users. They require generous amounts of storage space, powerful tools to extract and rescale image portions for particular uses, and GIS expertise to position the image correctly behind a map—not to mention the challenge of choosing among hundreds of very similar files, or of pinpointing which portion of an image covers a given location. Thus, despite the advent of CD-ROMs, fast processors, and improved mapping tools, effective use of orthophotos remains time-consuming and tedious—too much so for *ad hoc* use in many smaller government agencies and among a broad "non-specialist" audience. Putting these orthophotos on a networked server is attractive in that it would alleviate the burden of data storage and maintenance for these users. However, it would introduce limitations due to network bandwidth; and users would still need to choose orthophotos for a given location, create image excerpts of a practical size, and integrate them into a digital map. Therefore, this research effort focused on building server-side tools to facilitate finding and retrieving only the image portions needed, compressing them for efficient use of a limited bandwidth, and providing the header files as needed to integrate the image portion into GIS maps.

The orthophoto prototype also served as a pilot study for MassGIS, the GIS data center for the state of Massachusetts, which had begun to compile a statewide orthophoto series (some 2,000 grayscale images at 0.5 meter resolution, totaling 120 GB) and was exploring networked options for disseminating these to state and local agencies and the public.

Finally, I designed the project so as to test some of the ideas I'd been developing in my research on geographic information sharing. I wanted to balance the organizational understanding of geographic information infrastructures with an in-depth view of the technical components involved, their interconnection, and their interdependence.

b. Development stages

I developed the orthophoto prototype in collaboration with Joseph Ferreira, Jr., the project's Principal Investigator, and Thomas Grayson of MIT. I initially worked with Thomas Grayson on building metadata; he then focused on metadata search strategies, and I was principal architect for the graphical browser and server tools. Christian Jacqz and Carl Nylen at MassGIS provided orthophotos and documentation on CD-ROMs, and related technical support. Between early 1996 and early 1997, the MIT team built metadata files, processed the orthophotos, and performed all software development using the facilities of the Computer Resource Laboratory in MIT's Department of Urban Studies and Planning. The next several paragraphs (*i-v*) describe the stages in the development of the orthophoto browser; more details on the resulting product may be found in Section 4 below.

i. FGDC-compliant metadata for pilot orthophotos; begin manipulating images

This involved, first, learning the intricacies of the FGDC metadata standard. Early attempts to reduce it to a relational database form (for dual use as plain text display and searchable index) proved unsuccessful, and plain ASCII text became its canonical form. Verifying the metadata's accuracy also required repeated checking with MassGIS staff, reading photogrammetric project reports, etc. Simultaneously with this metadata work, MassGIS delivered its first set of (4) orthophotos, and I began experimenting with a freeware suite of graphics utilities known as *pbmplus*[6], to gauge their usefulness and performance for manipulating the large images in the MassGIS series.

ii. Preliminary Web-based interface to (tiled) orthophoto excerpts

At this point, designing the interactive browser's interface was my primary concern, and raised several questions about what capabilities to expect of the clients, what kind of networking would link the clients to the server, what users and uses to support, and how much server-side processing we could afford. These questions illustrate well the challenges that separate simply putting data online from providing a widely useful infrastructure for sharing geographic information, so it's worth spelling them out in some detail.

First, the HTTP and HTML standards, understood by any Web browser, provided good uniformity across a wide range of client software and hardware. (This uniformity has enabled many information services to be built in recent years that would otherwise have reached only a small audience.) However, because newer HTML elements might not be understood by all Web browsers, a frequent tension existed between providing useful features and including a wide audience. The varying capabilities of GIS systems provoked a similar tension when it came to integrating the browser with client-side GIS software. What to expect of client software was the subject of many lengthy conversations on the project team, and indeed the answers shifted considerably even within the project's time-frame.

As for networking capabilities, an important challenge was to use the network sparingly. Much of the Internet user base, *circa* 1996, had modem-quality access (14.4 to 28.8 kbits/s; magazines joked

---

[6] *Pbmplus* suite by Jef Poskanzer. Available online at <*http://www.acme.com*>.

about “the World Wide Wait”). To include a broad audience, it was important to send files of modest size, compressed wherever possible. Yet a growing portion of the audience did have high-bandwidth connectivity; and some important uses required large image sizes and highest-quality (uncompressed) images.

Third, although the project was clearly tied to geography and GIS by its funding and personnel, its intent was to extend the use of orthophoto data to a much wider audience (while still giving the GIS audience something worth their while). This raised another set of questions for the orthophoto interface: how well should the user know geography concepts, or the Boston-area geography? How would people find their way around the orthophotos? (For instance, would they know that the Charles River runs horizontally across the screen, separating Boston from Cambridge?) How to provide enough geographical information (coordinates, scale, etc.) for GIS users, without complicating the interface too much for the typical “Web surfer”? These questions, too, were the subject of many conversations.

Finally, several server-specific questions arose. For instance, given the quasi-continuous geographic space represented by the orthophotos, users should be able to choose excerpts in literally millions of locations and sizes—far more than could be prepared in advance. So, some (computationally costly) server-side image processing seemed necessary; yet response times needed to be low, a few seconds at most, for the browser to be properly interactive. So, how much real-time processing could the server afford to do in serving up these orthophotos? The answer depended on a combination of hardware performance and software efficiency. Multi-user access was another consideration: how many queries could the server handle at once before reaching memory or disk-space limits, or slowing down unacceptably? Managing the orthophoto service brought another set of questions. Any pre-processing of the images reduced the need to process them on-the-fly; but how many files, or how complex a data structure, were we willing to maintain? A good bit of trial-and-error was needed to resolve these questions.

Together, all these questions made the prototype effort much more than building something to suit a specification. Rather, the requirements were uncertain and changing, and choices along any of the above lines could "make or break" the usefulness of the data service. In particular, for this next phase of the project, because initial tests of the *pbmplus* utilities showed less-than-satisfactory performance on anything but the smallest images, I chose to defer all server-side image processing for later in the project. Instead, in April 1996 I built the first prototype around fixed, prefabricated image tiles (Figure 8-1), even though this would clearly limit the user's flexibility in navigating around the orthophotos, and would impose a significant file-management burden on the server. For this preliminary version of the browser, each orthophoto was pre-processed as follows: *(i)* cut each 8,000×8,000 pixel image into sixteen 2,000×2,000 excerpts; *(ii)* cut each of these into sixteen 500×500 tiles (Figure 8-1), storing these in compressed (JPEG) form; *(iii)* rescale the 8,000×8,000 and 2,000×2,000 images by a factor of 16 and 4, respectively, and store these in compressed form. This resulted in a series of 500×500 compressed image tiles, at three zoom levels, for each orthophoto. A script running on the server kept track of all these tiles: in response to a Web user's requests to zoom in and out, or to go north, south, east, or west, the script would select the tile that best fit the desired area, and present it as a "viewport" within an HTML page similar to Figure 8-5 below. Although each 8,000×8,000 master image is about 61 Mbytes, the compressed tiles displayed in the viewport were in the 80 kByte range, thus accessible even over dial-up.

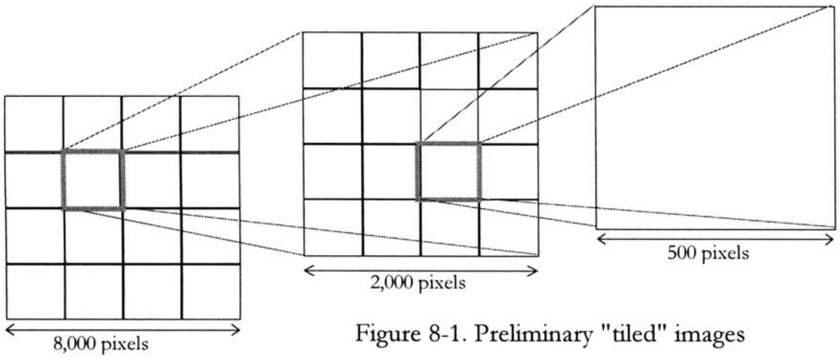


Figure 8-1. Preliminary "tiled" images

iii. Build a final orthophoto browser interface with custom image "snippets" and GIS headers

This preliminary interface was a convenient entrée into the images; its performance was entirely adequate on a Digital DECstation 5000; and it did use the network sparingly. But its fixed tiling scheme did not allow users to position their viewport very precisely, or to resize it to suit a variety of needs. Also, maintaining 273 ($1+16+16^2$) separate files for each orthophoto was an unacceptable maintenance burden on the server, especially when it came to including more image formats and tenfold more orthophotos. Furthermore, I found that I could obtain greatly improved performance in the *pbmplus* image-extraction utility (called *pnmcut*) through a small source-code change. So for this next stage, I looked into extracting images as needed from a single large master image.

In the initial "tiled" design of the orthophoto browser, the server-side script simply chose which one of many pre-existing image tiles to present to the user. For the second-generation orthophoto browser, the goal was to extract sub-images on the fly from a single 8,000×8,000-pixel master image for each orthophoto. This would allow users to position their viewport precisely, to set its size as large or small as they wished, and to choose any zoom level for viewing the images. It would also greatly simplify management of image files on the server. Finally, through in-line format translators, multiple image formats could be distributed from the same master image.

This required two server operations to be done on the fly: extracting a portion of a large master image, and rescaling the extract to a requested size. In July 1996, I modified the browser's script to call the newly streamlined *pnmcut*, and to "pipe" its output through a (JPEG) image compressor before displaying it in the viewport. (Details in Section 4 below.) Each request now created a new image excerpt, instead of just choosing a pre-existing tile: so users could zoom in or out on a precise spot, re-center the map on a precise location, and size the viewport as needed. This image excerpt was not stored anywhere on the server, but was piped to the user as a stream of bytes preceded by an image header. By simply inserting various image-format translators into the pipe, it also became possible to serve up the same image-excerpt in a variety of formats, all extracted on the fly from a few large master images.

I had hoped to use another *pbmplus* utility, called *pnmscale*, to rescale image excerpts on the fly. Combined with the *pnmcut* utility above, this would have given the user a continuous set of zoom levels, all extracted from a single 8,000×8,000 image. However, rescaling images is a by nature a compute-intensive process, and the *pnmscale* utility reflected this in its slow performance. To maintain acceptable response times, therefore, my fallback was to maintain three master files for each orthophoto: one at each of three fixed zoom levels (8,000×8,000, 2,000×2,000, and 500×500). This provided somewhat jarring jumps (1:4 or 4:1) when zooming in or out; but it maintained good performance, and the file-management burden was still quite acceptable (only 3 files, instead of 273, for each orthophoto; the two rescaled versions only occupied 7% additional disk space).

I used the same "piped" approach to provide the geographic header files needed to incorporate each image excerpt into a desktop GIS package on the client machine. The resulting interface is shown in Figures 8-5 and 8-6 below.

Of course, all this interactive functionality required the server to do much more work than before in response to each of the user's mouse-clicks. To ensure continued rapid response with many simultaneous users, in August 1996 we moved the orthophoto service from the DECstation 5000 to a dual-processor Sun Microsystems SPARCserver 1000E, tuned for fast computation and network data services. Interestingly, the "piped" server-side processing introduced no delay in the response time (the client starts receiving bytes immediately); instead it limited the output rate to about 90 kBytes/s. This is about half of T1 bandwidth (1.544 Mbits/s), still faster than most Internet lines, as of mid-1997. (In other words, in most cases, the server still has to wait for the client to catch up, not the other way around.) As typical bandwidths increase in coming years, however, this could become a bottleneck and may require faster hardware or software, different compression schemes, or smart caching on the client or server.

Finally, I made a small change to the script, whereby parameters specifying the image excerpt appeared in the URL[7] used to call the script (previously these parameters were "posted" invisibly to the script). This allowed users to bookmark particular orthophoto views in their browsers for future reference, and to embed the dynamically created images into their own Web pages. These images would not be stored anywhere (client or server); rather, displaying such a page onscreen would trigger the orthophoto browser script via the Web. As run-of-the-mill software tools become more integrated with Web services, these dynamically-generated images may well find their way into online documents, spreadsheets, or digital maps. This would free the orthophoto service from the confines of a Web browser, and weave it into all kinds of everyday work.

---

[7] A URL, or Uniform Resource Locator, is the unique address of any file or byte-stream on the Web.

iv. Unveil full-scale service to a wide audience; examine management and replication issues.

As the second-generation orthophoto browser was nearing completion, additional MassGIS orthophotos became available, ultimately reaching a total of 118 orthophotos (7 GB, with an additional 500 MB for the two rescaled versions of each orthophoto). This prompted the design of scripts which kept the data-maintenance effort manageable. In late September 1996, we announced the orthophoto service on a variety of email lists: its use quickly grew to about 14,000 "hits" per day, and has persisted at about 1,500 hits per day, from a broad cross-section of Internet users (Figures 8-2 and 8-3).

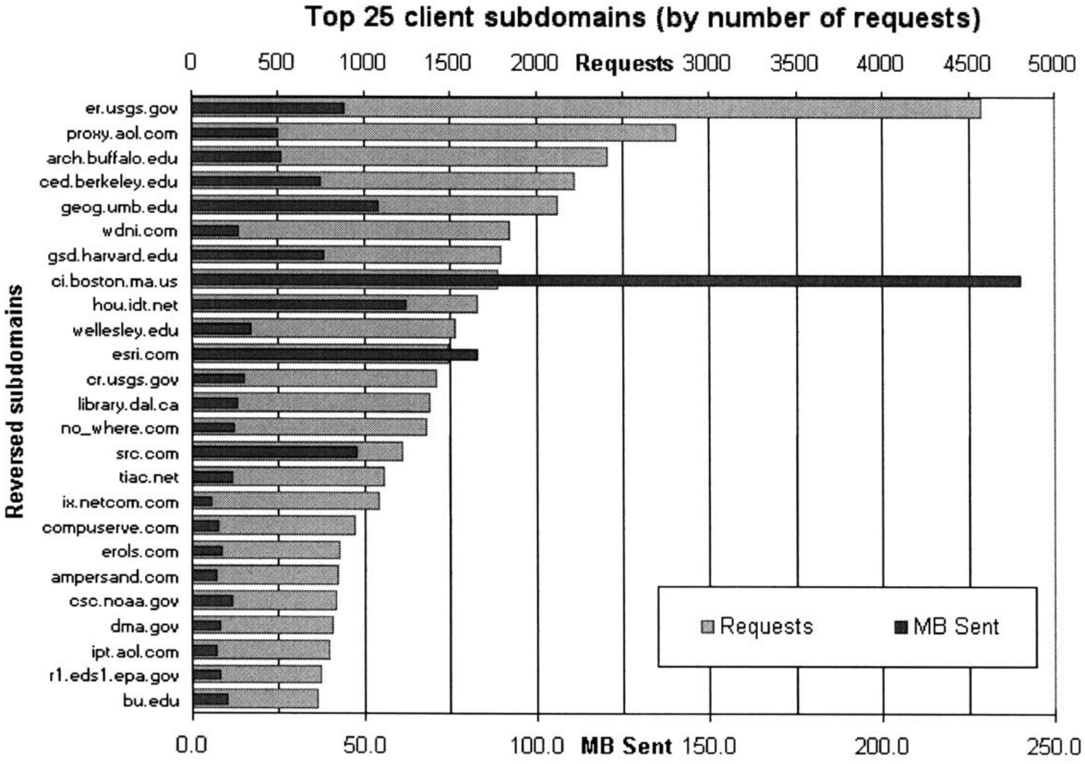


Figure 8-2. Internet subdomains using the orthophoto browser, Sept. 1996-March 1997[8]

---

[8] Admittedly, this figure raises more questions than it answers. For instance, it describes total requests (a single Web page may produce many separate requests for inline graphics); it has a large portion of "unresolved" addresses which would need to be tracked down; and some sites such as proxy.aol.com may correspond to hundreds or thousands of users. Nonetheless, it shows the variety of usage and the discrepancy between queries that retrieve large image files (lots of MB per request) and those that retrieve simple text (lots of requests per MB).

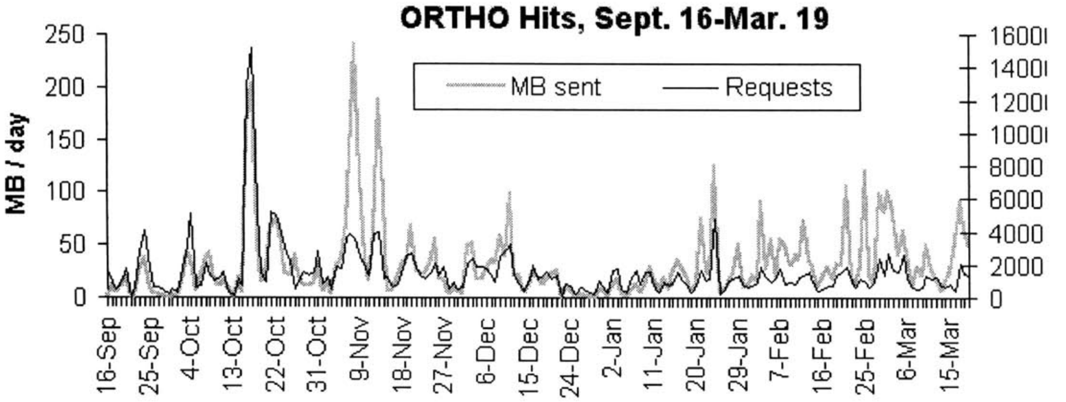


Figure 8-3. Orthophoto browser usage, Sept. 1996-March 1997

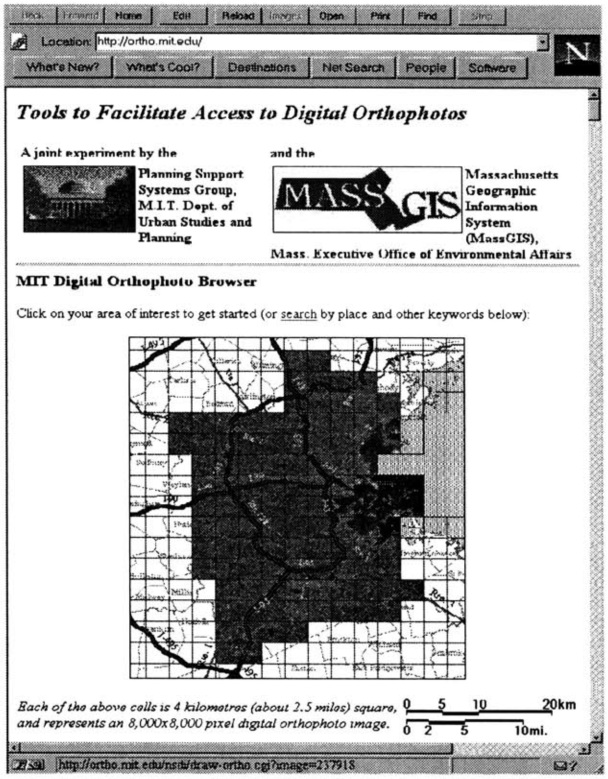
I then focused on the replication of the orthophoto browser at other sites. First, by assembling a similar browser interface for color orthophotos of the Washington, DC area, I confirmed that the scripts I'd written could be fairly easily adapted to other orthophoto series. Also, in keeping with its decentralized nature, the NSDI Clearinghouse seemed most likely to grow in an "organic" fashion, as Internet-accessible sites learned to become Clearinghouse nodes. To this end, I wrote a tutorial documenting the various components of the orthophoto service, and explaining how to set them up for orthophotos at other sites.

v. Discussion of development stages

These stages of development are more than just a "war story": they illustrate a more general set of tradeoffs to be balanced in designing components of geographic information infrastructures, given that available GIS and networking technologies are in rapid transition. First, despite an otherwise sound design, server and network performance can make or break a given strategy. For instance, the new design in stage 3 above wasn't feasible without a few changes to the *pnmcut* image-processing utility. Also, moving to a fast server and using image-compression schemes allowed useful interactivity in the final version of the browser: without them, the browser would have remained an interesting novelty, but too slow for any practical use. Second, these stages illustrate the multiplicity of answers as to how software systems can interact, or what capabilities can be counted on when designing either GIS systems or networked data services. For instance, a primary thrust of this project was designing an interactive HTML interface; but as future GIS software provides increased network integration, the orthophoto service will probably be expected to handle a variety of geographic queries. These and other choices took into account current technological conditions (*circa* 1996), and they may or may not continue to make sense as these conditions inevitably change. This highlights the importance of building systems like this one in a modular, or "layered," fashion, so that components can be swapped in or out as needed; also, if these components are already available then a project like this one can focus its efforts on "bundling" them in creative and flexible ways rather than building them from scratch.

4. Product functionality

The final orthophoto prototype, accessible at the URL <*http://ortho.mit.edu*> (Figure 8-4), provides simple interactive access to digital orthophotos. Four aspects of the prototype distinguish it from more conventional approaches to distributing geographic information: its interactive geographical interface, its links to client-side GIS software tools, its dynamic creation of custom data products from a single set of master data, and its links between the data and searchable metadata. It also helps to bridge the barrier of geographic scale by providing a detailed geo-referenced "canvas" for creating line (vector) maps of many kinds. This makes it a promising example of the style of technology needed to build practical, high-impact infrastructures for geographic information.



Tools to Facilitate Access to Digital Orthophotos

A joint experiment by the Planning Support Systems Group, M.I.T. Dept. of Urban Studies and Planning and the Massachusetts Geographic Information System (MassGIS), Mass. Executive Office of Environmental Affairs

MIT Digital Orthophoto Browser

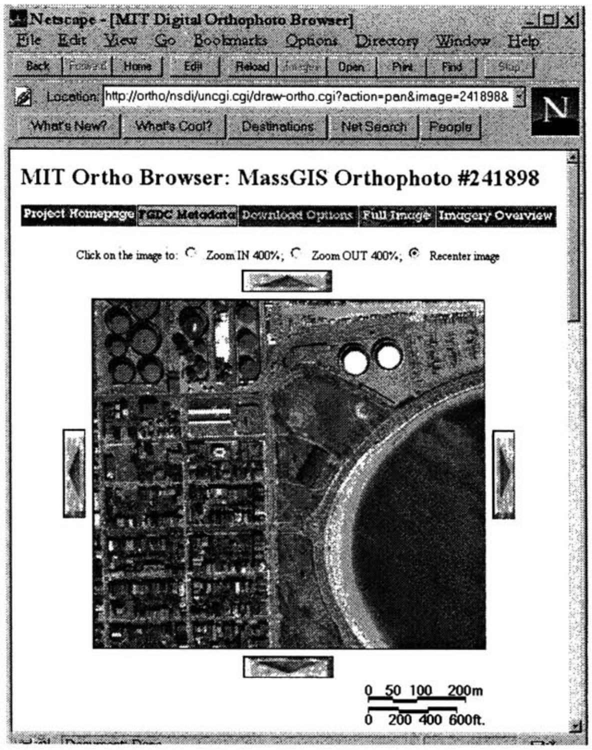
Click on your area of interest to get started (or search by place and other keywords below):

Each of the above cells is 4 kilometres (about 2.5 miles) square, and represents an 8,000x8,000 pixel digital orthophoto image.

Figure 8-4 Orthophoto project homepage

a. Browser overview

The image browser, depicted in Figure 8-5, is the primary access point for most users. It displays a viewport showing a portion of an orthophoto, with controls that allow the user to pan across an image and to adjacent images and to zoom in or out on the image. Each time the user requests a new view, the server extracts a custom image from a master image, and sends it to the user in compressed (JPEG) form for viewing. At the top of the browser page are buttons that link to the project's main page, to metadata on this image, and to imagery overviews. The viewport is oriented with North pointing up, and a dynamic scalebar tracks the scale at each zoom level. (Not shown in the figure are controls for resizing the viewport, from postage-stamp size to more than full-screen.)



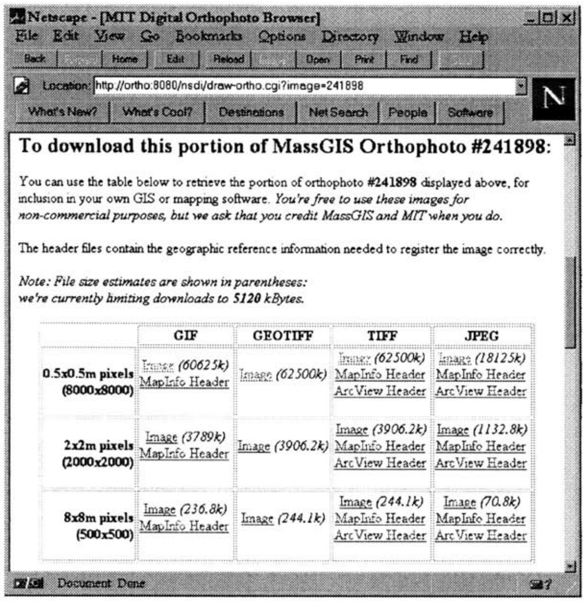
http://ortho/nsdi/uncgi.cgi/draw-ortho.cgi?action=pan&image=241898&

MIT Ortho Browser: MassGIS Orthophoto #241898

Click on the image to: ○ Zoom IN 400%; ○ Zoom OUT 400%; ● Recenter image

Figure 8-5. Orthophoto browser main page

In another part of the interface (Figure 8-6), the viewport image can be downloaded in several file formats (GIF, GeoTIFF, TIFF, and JPEG) and resolution levels (½ , 2, and 8 meter pixels). To prevent abuse of the service, image file sizes are estimated in advance, and requests that exceed a (fairly generous) limit are denied. Users can also obtain geographic header files for the image, so as to overlay it with digital maps using GIS software such as ArcView® or MapInfo®.



http://ortho:8080/nsdi/draw-ortho.cgi?image=241898

To download this portion of MassGIS Orthophoto #241898:

You can use the table below to retrieve the portion of orthophoto **#241898** displayed above, for inclusion in your own GIS or mapping software. *You're free to use these images for non-commercial purposes, but we ask that you credit MassGIS and MIT when you do.*

The header files contain the geographic reference information needed to register the image correctly.

Note: File size estimates are shown in parentheses:
we're currently limiting downloads to ***5120*** *kBytes.*

	GIF	GEOTIFF	TIFF	JPEG
0.5x0.5m pixels (8000x8000)	Image *(60625k)* MapInfo Header	Image *(62500k)*	Image *(62500k)* MapInfo Header ArcView Header	Image *(18125k)* MapInfo Header ArcView Header
2x2m pixels (2000x2000)	Image *(3789k)* MapInfo Header	Image *(3906.2k)*	Image *(3906.2k)* MapInfo Header ArcView Header	Image *(1132.8k)* MapInfo Header ArcView Header
8x8m pixels (500x500)	Image *(236.8k)* MapInfo Header	Image *(244.1k)*	Image *(244.1k)* MapInfo Header ArcView Header	Image *(70.8k)* MapInfo Header ArcView Header

Figure 8-6. Choosing image and header formats

Behind the interface is a script on the server (draw-ortho.cgi), written in the *perl* language (Wall et al., 1996), that receives parameters from the user via the HTTP[9] Common Gateway Interface (CGI). These parameters are embedded in a lengthy URL like the following:

---

[9] HTTP is the Hypertext Transfer Protocol, the machine-to-machine language of the World Wide Web.

```
http://ortho.mit.edu/nsdi/draw-ortho.cgi?action=zoomin&dwidth=600&dheight=400-
&x=252&y=189&image=237898&width=400&height=300&middlex=250-
&middley=250&zoom_level=4
```

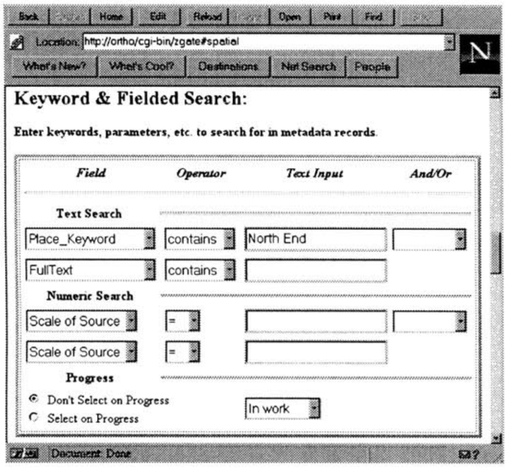
The parameters include, first, an "action" to be performed ("zoom in," "pan," "go north," etc.), along with the desired viewport size and where needed, the coordinates of the user's mouse-click. These are followed by several state variables passed between client and server, that keep track of the current image name, viewport size and coordinates, and zoom level.

Based on these variables, the `draw-ortho.cgi` script computes the coordinates of a new image excerpt, and then calls another script (`stdout.cgi`) which runs several software utilities in the *pbmplus* freeware suite to extract the sub-image from the full master image, reformat it to the desired image format, and pass it back to the user via Unix's "standard output" byte-stream.

Two more scripts (`mi-header.cgi`, `av-header.cgi`), when activated by the user, compute the geographic location of the image and return it as a header file formatted for MapInfo or ArcView.

b. Links to metadata

Each image accessible through the orthophoto browser is associated with a set of metadata elements that describe the image according to long list of criteria set forth in the Federal Geographic Data Committee's (FGDC) Metadata Content Standard. This is useful in two ways: first, while looking at an image in the orthophoto browser, a user can click on the "FGDC Metadata" button (cf. Figure 8-5) to review the author, date, precision, and processing history (etc.) of the image. Second, before perusing the images graphically, users can find out which ones to look at by querying the orthophoto metadata as a library catalog. This is done through a simple form interface which lets users search on keywords in specific fields (place, author, etc.), as well as on location (a latitude-longitude bounding box) or the date of data capture or publication. The forms interface is shown below:



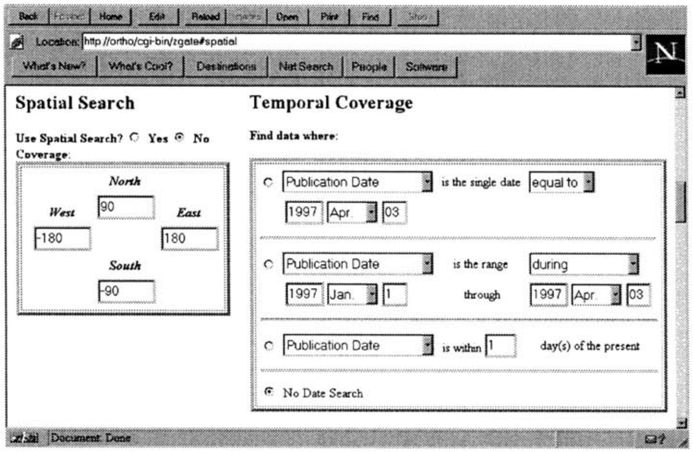
http://ortho/cgi-bin/zgate#spatial

Keyword & Fielded Search:

Enter keywords, parameters, etc. to search for in metadata records.

Field	***Operator***	***Text Input***	***And/Or***
Text Search			
Place_Keyword	contains	North End	
FullText	contains		
Numeric Search			
Scale of Source	=		
Scale of Source	=		
Progress			
● Don't Select on Progress ○ Select on Progress		In work	

Figure 8-7. Metadata: keyword search



http://ortho/cgi-bin/zgate#spatial

Spatial Search

Use Spatial Search? ○ **Yes** ● **No**

Coverage:

North 90
West -180
East 180
South -90

Temporal Coverage

Find data where:

○ Publication Date is the single date equal to 1997 Apr. 03

○ Publication Date is the range during 1997 Jan. 1 through 1997 Apr. 03

○ Publication Date is within 1 day(s) of the present

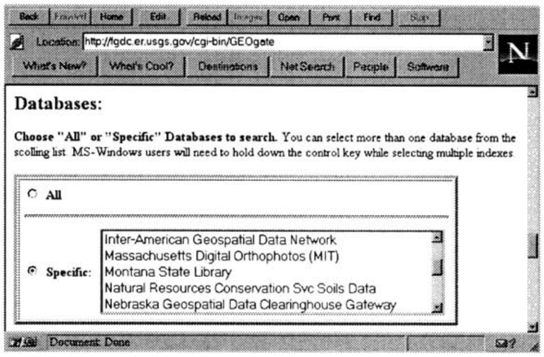
● No Date Search

Figure 8-8. Metadata: spatial and temporal search

For instance, a user searching for images covering the North End (a neighborhood in Boston) can request that string in the "Place-Keyword" field. Upon receiving this query, the server returns a set of "titles," hypertext links to the images in question. The user can click on one of these links to see the full set of metadata—of which one element is a hypertext link to the image itself, viewed through the orthophoto browser.

Besides this interface, these metadata are also accessible to any networked client via the Internet's Z39.50 protocol (Kahle, 1991). Because they also conform to the FGDC's Content Standard, users can search this and many other metadata sites simultaneously to find and review items of interest regardless of their location on the network.

The collection of standard-content metadata accessible via the Z39.50 protocol is known as the National Geospatial Data Clearinghouse, which users query through multi-site search services like that at the FGDC's Clearinghouse Web site (*www.fgdc.gov*). The FGDC's query form is identical to the one at *ortho.mit.edu*, but with an additional section, shown in Figure 8-9, for multi-site



http://fgdc.er.usgs.gov/cgi-bin/GEOgate

Databases:

Choose "All" or "Specific" Databases to search. You can select more than one database from the scolling list. MS-Windows users will need to hold down the control key while selecting multiple indexes.

○ **All**

● **Specific:**
Inter-American Geospatial Data Network
Massachusetts Digital Orthophotos (MIT)
Montana State Library
Natural Resources Conservation Svc Soils Data
Nebraska Geospatial Data Clearinghouse Gateway

Figure 8-9. Multi-site query for the National Geospatial Data Clearinghouse searches. Through this form and others like it, the Clearinghouse lets users find the orthophoto browser without knowing in advance its address or even its existence.

5. Evaluation

The orthophoto prototype can be evaluated in several ways: its fulfillment of the CCAP project terms, its demonstrated value to researchers, reports of its use, and its performance along the yardsticks set forth for the cases.

a. Fulfillment of project objectives

First, the orthophoto browser fulfilled the terms of the FGDC project by providing a Clearinghouse node with Z39.50 searchable metadata on the orthophotos. It also provided access to the orthophotos themselves, with reusable tools for others to replicate similar services elsewhere. A final report is in progress.

b. Demonstrated value to the research community

Second, the prototype provided opportunities to reflect on how emerging technologies, or current tools assembled in new ways, can impact geographic information sharing. The prototype has already led to several invited conference presentations as well, including GIS/LIS, Auto Carto, and the New England and Pacific Northwest Chapters of URISA.

c. Usage reports

Another way of evaluating the prototype is a purely pragmatic one: Was it useful? The charts in Figure 8-2 and 8-3 tell part of the story, but another measure of its usefulness came in the form of anecdotes, mostly by electronic mail. In Boston's Public Facilities Department, a data manager finds that the orthophotos provide quick, easily understood views of building footprints, vacancies, and vegetation patterns in various urban neighborhoods: these are useful at community meetings in conveying the scope and impact of the City's plans. At the Boston Redevelopment Authority, members of the research department use the orthophotos to correct and update their existing street maps: the pictures help them to realign their street lines more precisely and to weed out lines that correspond only to rights-of-way through existing blocks or parcels. In November, 1996, someone from Boston City Hall (*ci.boston.ma.us*) downloaded the entire collection of orthophotos at the intermediate zoom level (2m pixels), leaving behind noticeable spikes on the charts in Figures 8-2 and 8-3 above. We also received comments from GIS and other researchers and educators, mapping firms, students, a major logging company, federal and state agencies, and many others.

d. Performance along the case-study criteria

One last way of evaluating the orthophoto prototype is to use the same criteria as those used in the case studies. As before, this evaluation is in two parts: the user's perspective, focused on the data being shared and its use, and the architect's perspective, focused on maintaining and growing the mechanisms of information sharing.

First, the timeliness and concurrency of the prototype are based on the update frequency of the orthophotos (every few years at best) and the rate of change of the urban landscape they represent. The locational precision of the images (0.5m pixels) and their accuracy (90% of pixels less than 2.5 m off) are enough to suit a wide range of mapping purposes, such as the Boston Redevelopment Authority's use of them as a backdrop "canvas" for geo-referencing vector maps. The semantic precision and accuracy of the images are related to their radiometric resolution (here, 220 different shades of gray) as well as factors such as cloud cover or sun and shadow angles. Unlike conventional vector maps or tabular databases, the individual pixels contain no meaning as to land use, hydrology, or roads—let alone property lines or state boundaries. However, visual interpretation of the patterns of light and dark easily fills in semantic content. The precision or accuracy of this visual interpretation varies widely across the image: for instance, water bodies are easily identified by their dark color, but a grassy field may have nearly the same grayscale luminance as an asphalt road.[10] Here again, this is enough detail for the two example uses: it enables people at community meetings to see clearly what lies behind the lines of a property or zoning map; and it provides a clear view of even the smallest city blocks for maintainers of street files. Finally, the browser scores high on the encapsulation criterion, given that it puts image data files into any of several common

---

[10] Color orthophotos, which store multiple luminance values at each pixel, would allow greatly enhanced visual interpretation, as well as automated statistical classification of pixel "clusters."

file formats and also prepares the necessary header files for loading the images into GIS/mapping packages.

Second, taking the "architect's" view, reciprocity is one area in which the orthophoto prototype does not score very high. As a one-way data distribution mechanism, it provides only informal mechanisms (such as unstructured electronic mail) for "upstream" data sharing. An interesting extension to such a browser would be a facility for users to send their own orthophotos or maps for display in the browser viewport—or even to send lists of coordinates or street addresses for others to overlay them atop the orthophotos. The HTML ENCTYPE tag may be helpful in providing this capability in a future version. Next, scalability is one criterion along which the orthophoto browser measures up rather well. It withstands the load of several thousand data requests per day with no noticeable delays; its open access accommodates any client host; it runs on a single server at present, but its different tasks could be easily split among several (e.g. one just for images, or even one for each type of image). Finally, by producing images and header files in multiple formats, the prototype attains a high degree of non-intrusiveness by accommodating, from a single dataset, a variety of user requirements. Furthermore, it doesn't impose free and open access: the prototype restricts access based on file size, but any number of restrictions are possible, based on time-of-day, client domain, image quality, server load level, etc. The service could be restricted to subscriber sites, or could even charge users pennies per transaction.

6. Implications: what's different about the orthophoto browser?

What does the orthophoto bring to those who will build the National Spatial Data Infrastructure? To the agencies I observed in my case studies? To other, similar contexts? To answer these questions, I will first summarize what's different about the orthophoto browser compared with more usual methods for sharing geographic information.

By instantiating many of the technical recommendations arising from the case studies, the orthophoto browser shows what different technological choices may be brought to bear. In particular, it provides a widely accessible "just-in-time" geographic data service, it supports access through metadata, it enables interoperability between different software systems, and it suggests promising directions towards scale-free geographic referencing.

a. A widely accessible, "just-in-time" data service

Unlike more conventional mechanisms for sharing information that rely on putting fixed data files online, the orthophoto prototype prepares custom data products as needed. (Note that this required tuning the components of the server, to account for a wide range of client software and network connectivity.) This alone could play a key role in facilitating the integration of data maintenance and electronic distribution in the case-study contexts. First, thanks to the orthophoto browser's "just-in-time" architecture, maintaining the orthophotos for both internal and external use remains easy, with no proliferation of temporary files or multiple uncoordinated copies of the same data. Second, access methods are appropriate for a wide variety of users: novices and GIS experts alike, whether in the next room or across the Internet, can get to the data they need quickly, with little need for assistance. Compared with a conventional online archive such as an *ftp* server, or a data center selling preset or special-order bundles of images on CD-ROMs, the orthophoto browser requires a lot less work on the part of either the data provider or the user to make the data useful. Third, the data can be corrected or updated even after the service is launched, so careful documentation of information may impose less of a delay. In summary, the prototype turns a large dataset into an easily used, low-maintenance data service. Gateways like this one could do much to merge the often separate activities of internal data management and external data distribution as seen in the cases, as well as the dichotomy between expert and novice uses.

b. Information integration through metadata

This service addresses the need for metadata to search for and interpret geographic information in a few simple ways: it provides header files for incorporating the images into several GIS software systems; and it provides a full FGDC metadata listing, which can be searched via the Z39.50 protocol, or reviewed in detail to determine the fitness of the orthophotos for a particular use. The metadata's standard content also offers the promise of direct interaction between the orthophoto service and future client-side software systems.

c. Interoperability between software systems

Third, the prototype helps to shift the focus of geographic information sharing from compatibility between datasets to interoperability between software systems. It does this in several ways: embed orthophoto URLs into other Web pages (and, soon, documents, spreadsheets, and maps); the Z39.50 protocol which lets users query the orthophoto metadata simultaneously with other such services across the network; and the provision for several different client-side GIS systems. Several additional kinds of interoperability could be built: the ability to specify locations by latitude and longitude or even by street address; and, more generally, a standard language for querying the orthophoto server. As such, the prototype heralds the unbundling of today's large software applications into a networked "fabric" of independent but interoperable data services. This of course has organizational ramifications, which will be explored further in the next chapter

d. Towards a scale-free referencing system

Finally, the Pacific Northwest's experience with River Reach files in particular highlights the challenges of building a consistent geographic reference system, and then evolving it to suit increasingly demanding needs and an increasingly complex set of technology choices. The choice in that case was to store properties of the river network as a table of properties tied to river reaches, and cross-reference these river-reach codes to those in the various states. A newer approach to this problem, based on "dynamic segmentation," stores "events" tied to a linear "river mile" referencing system and computed their geographic locations only when displaying them on a map. However, the same river may measure twice as long at a 1:24,000 scale as it does at 1:100,000. Thus, upgrading an event table to a more detailed map scale is not an easy process: it requires computing geographic startpoint and endpoint coordinates for events using the less detailed river features, and then (through a variety of geographic approximations) building a new event table for use with the more detailed streams.

Putting easy access to orthophotos at everyone's fingertips may offer a more accurate alternative: if field workers were to use highly precise orthophotos as a backdrop to enter observations and measurements, their "events" could be tagged with geographic coordinates (e.g. latitude/longitude), rather than linear coordinates, and they would be less dependent on a particular map representation. As the need for geographic detail and precision grows, it wouldn't be hard to swap in more precise orthophotos where needed; or conversely to keep only low-resolution ones in areas with few features of interest. This approach may not be completely scale-free (the pixels do have some size); but by providing precise locations instead of scale-dependent feature geometry, significant gains in precision and accuracy may be possible. The orthophoto browser isn't the last word on the issue; rather it highlights ways in which a few improvements in technology and system performance can make a continuing problem marginally easier to address.

7. Implications for the National Spatial Data Infrastructure

Due to its unique features, the orthophoto prototype suggests several useful directions for the National Spatial Data Infrastructure (NSDI)'s strategies: it explores ways to grow the NSDI Clearinghouse into a source for integrated geographic data and metadata; it shifts the NSDI standards focus onto software interfaces and query languages, and it shows how networked access to digital orthophotos might play an important role in the NSDI's Framework.

a. Growing the NSDI Clearinghouse beyond metadata

The orthophoto browser shows how the Clearinghouse might grow to provide both geographic data and metadata in an integrated fashion. Its detailed, standardized metadata allows users to evaluate the image data and its appropriateness for their use; also, users trying to locate relevant geographic data for a project can also search this site in parallel with other metadata sites. Metadata records also link directly to the orthophotos themselves, within a multi-purpose interface. Of course, one key reason for the Clearinghouse's focus on metadata is that not everyone wants to give away their geographic data files: this prototype's very simple mechanism for access control (based on file size), suggests any number of mechanisms for restricting access to certain sites or users, allowing only sample access, or requiring lump-sum or per-byte payments.

b. Shifting the focus of NSDI standards

By preparing images and header files in a variety of formats, the orthophoto browser begins to delve into the larger problem of GIS interoperability. Greater integration with client-side GIS tools would require two kinds of standards: a commonly accepted way of specifying geographic queries, and a way to send queries from these tools to a Web browser or Web server. Although it's possible to create special-purpose software to integrate a few special cases (such as one or two different GIS packages), more general solutions await agreed-upon standards for specifying geographic data entities and requests. Whereas many standards efforts thus far have emphasized *data* standards such as file formats, the orthophoto prototype's small foray into interoperability suggests that *functional* standards, such as query languages and software interfaces, may recast the NSDI as a link among heterogeneous geographic software components.

c. Networked orthophoto services in the NSDI Framework

Finally, the prototype's use of high-resolution raster data provides a novel form of geo-referencing, allowing users to enrich their maps visually and to update their line datasets. This in itself is an important contribution to the NSDI Framework; but in addition, this prototype suggests that the Framework may consist not just of data files but of active data services that can prepare custom data products for each user and function within a limited network bandwidth.

8. Implications in the case-study contexts

a. More-than-incremental change

Finally, a key finding from the cases was the difficulty in moving from a temporary "scaffolding" to a more long-term distributed infrastructure. On both the technological and organizational fronts, this may require forgoing incremental changes in favor of more radical changes: that is, "leapfrogging" to completely new ways of doing things instead of building on what's known. Compared with conventional vector GIS, or indeed tabular databases, the highly precise geographic referencing provided by these orthophotos opens up completely new ways to tie information from many different sources to spatial locations within a single map. Furthermore, the move to networked services in information sharing may effect radical shifts in the management of information systems, from guarding a comprehensive set of data and software applications, to finding reliable sources of information and processing services and forging interoperable links to these sources. Some will prefer a more secure data management plan, where an organization maintains all the information it counts on for its work. And especially in the area of geographic information, reliable, interoperable data services are still very few in number. So before these radical, non-incremental changes can occur, functional standards may need to take hold to support interoperability, and unbundled geographic processing services may need to become more widespread and reliable. So despite the need for non-incremental change expressed above, organizational change and technology may need to evolve in parallel and thus more incrementally. Chapter 9 explores standards, services, and change patterns in more depth.

b. Who pays? Distributing responsibilities, costs, and benefits

Finally, one implicit theme underlies the case studies: "why bother building infrastructures to share information"? The costs of building infrastructure systems are usually immediately tangible, considerable, and unevenly distributed. Yet their benefits are often distant and uncertain, and often accrue to many more users than those that paid for their development. The next chapter discusses a variety of ways to resolve these tensions, either through government intervention (subsidies, seed money, standards) or by providing some return on investment (fees, prestige, barter, etc.). Related to this is the need to share the data maintenance responsibilities in a reliable fashion among many players.

In short, the orthophoto prototype provided a concrete setting for surfacing the interplay of technological and organizational aspects of building relevant spatial data infrastructures. Whereas the case studies of Chapters 4-7 shaped views of what choices are worth making in technological design, the prototype experience helped to interpret the case experience and ways of extrapolating issues towards the future. Chapter 9 explores these issues in greater detail, as they apply to technology choices, organizational structures, and policy priorities most conducuve too building effective geographic information infrastructures.

Chapter 9: Some implications for technology, organizations, and policy

1. Overview

a. Motivation

What does it take to share geographic information? That question back in Chapter 1 was motivated by the observation that it's rare for planners or public managers to share their working information with each other across organizational boundaries. It's rare even in environmental planning and policy, where it would seem especially beneficial due to key ecological relationships across jurisdictions, industries, and government hierarchies. It's rare despite the nature of geographic information, whose high cost and potential for widespread re-use should discourage duplication of efforts; and even though the benefits of combining multiple geographic datasets for a given location should discourage isolated, parochial approaches. The difficulty seemed to be adjusting both technological and organizational systems to accommodate shared geographic information. One likely solution was an infrastructure for sharing geographic information—that is, a package of organizational and technological choices geared towards helping organizations make use of each other's information resources in their work. Learning how to design and grow such infrastructures seemed key to effectively sharing geographic information for environmental planning and policy. The resulting research effort had two primary thrusts: case studies and a prototype data service.

b. Findings from the case studies

The case studies examined three real-life examples of geographic information sharing infrastructures linking environmental agencies, and compared their technological and organizational characteristics and their growth patterns over time. The cases highlighted the importance, in launching geographic information infrastructures, of a convergence between shared norms, resources, and people to articulate these norms and leverage the resources. Once launched, the cases showed, infrastructures risked getting stuck at an intermediate "scaffolding" stage of development, with few tangible impacts on planning and policy. At these and other choice points, they needed someone to integrate many views of infrastructures and information sharing, so as to grow the organizational and technological complexity needed to affect real decisions and to be sustained over the long term. Indeed, given a complex and rapidly changing context, the choice of a growth path and the unfolding of other decisions over time seemed more important in the cases than a set of initial factors or an *a priori* blueprint. Nonetheless, a *laissez-faire* approach was inadequate: some evolving standard (a geographic reference system, or functional standards such as metadata or queries) was important to build convergence among participants. Finally, deeper-than-expected organizational changes seemed necessary to capitalize on a new "data services" model of information sharing, in which data management and communications were merged, and interdependence and teamwork governed a complex "ecosystem" of government agencies. Yet because organizational and technological changes were interdependent, both kinds of change were more likely incremental than radical.

c. Findings from the orthophoto prototype

The second thrust of the research put into practice some ideas about data services for geographic information sharing. Built from simple, freely-available software components, the orthophoto browser provided an efficient, customized online service for geographic information, through "just-in-time" extraction and compression of image snippets, along with header files for integration with local mapping software. This opened up use of orthophoto data to a much wider audience, in a way that encouraged convergence of geographic data among different sources. This service suggests an expanded conceptualization of the National Spatial Data Infrastructure (NSDI), as a collection of networked services and not just static datasets. It also foretells a need to shift the NSDI's standards focus, and that of other similar information infrastructures, from traditional data standards to newer functional standards that govern the interaction between information systems. The experience of building the browser also highlighted the many key design choices involved, and their ephemeral nature given the pace of technological change and the loose coupling between clients and servers. In particular, it seemed that designers of geographic data services would often face the following challenges: providing useful features *vs.* reaching a wide audience; building for a widely diverse set of users and uses; and tuning the service for current hardware and widely variable networking and client software. The resulting prototype also provided a tangible view of the organizational changes implied for the three cases and other similar contexts, as agencies using data services redistribute responsibilities for information collection and management.

d. Synthesis of the two studies

Together, these two research efforts—case studies and orthophoto prototype—provided insights that were both grounded in real-world organizations, and sensitive to coming technological changes, to help public agencies reap the promise of shared geographic information for ecosystem planning and policy. This chapter draws on both of these research efforts to articulate strategic choices of technology, organizations, and policy for geographic information infrastructures.

Section 2 below shows that in building geographic information infrastructures, standards are strategic choices that define the nature, scope, and effectiveness of the infrastructure as a whole. Particularly in the geographical arena, choosing what to standardize is a complex, multidimensional decision. On the organizational front, Section 3 suggests that new structures and relationships are both necessary and likely, and that agencies face the dual challenge of redistributing information responsibilities and balancing incremental *vs.* radical change. Section 4 shows several ways in which government policy can foster open, flexible infrastructures; and several ways that these infrastructures may enhance environmental policy. The last section sketches important questions for further research.

2. Technology implications: choosing strategic standards

In reflecting on the case studies and the prototype, one significant theme emerges: the importance, and the challenge, of setting standards for sharing geographic information. In an interorganizational context, a *laissez-faire* approach is attractive: let everyone do as they please and translate things as needed, with no change required of infrastructure participants. However, this approach has its limitations. In the Great Lakes and Gulf of Maine cases, for instance, the infrastructure imposed no standards other than World Wide Web connectivity: but the Great Lakes Information Network had trouble tapping into the region's geographic datasets in any systematic way, and the Gulf of Maine EDIMS never did spawn the additional data sources that had been hoped for. In contrast, in the Pacific Northwest, the choice to use a standard geographic reference for the Northwest Environmental Database, and standard criteria for river assessment, was essential to bringing together the states' river information for mutual benefit.

These examples suggest that although setting standards may require some adjustments on the part of an infrastructure's participants, a little standardizing can go a long way: the benefit of collaboration based on shared information may far outweigh the cost of well-chosen standards. Furthermore, standardization may be less onerous than translation for data that are complex or require approximation and interpretation—for instance, when rasterizing a vector map, conflating a network to a different map scale, or reconciling land-use codes. Therefore, counting on translators to minimize the need for standards isn't always practical, especially in the case of geographic or scientific data which is an always-imperfect representation of a physical reality. Instead, it's more important to choose standards strategically in light of the goals and constituents of the infrastructure. In this sense, to employ structuration lingo, a standard is not just a *rule* restricting the kinds of data, interfaces, or languages that the infrastructure will support, but also a *resource* that enables certain kinds of joint work through information sharing. (Admittedly, this isn't a terribly new concept: Dertouzos (1997b) quotes Simon (1981) to suggest that the makers of interconnected computer systems need to define the "laws" that will make them useful and encourage their orderly growth.)

The question becomes, then, not whether to standardize, but what part of the information sharing to standardize: what standards are likely to give the biggest payoff (in terms of collaborative work based on shared information) in relation to their cost (restructuring databases, translating keywords, etc.) for what a certain set of people want to accomplish. Depending on the purpose of the infrastructure, and its intended participants, it may be useful to standardize data structures, for instance, or units of measure, human or machine languages, or data collection methods. Data services such as Chapter 8's orthophoto browser may shift the focus of standards to metadata elements, query languages, object and method specifications, or shared basemaps for geographic reference. An infrastructure may employ several of these standards at once, in a layered fashion that supports multiple uses and allows standards to be changed independently of each other (Solomon and Rutkowski, 1992). For instance, regardless of whether the Pacific Northwest River Reach are in Arc/Info or Intergraph format (or distributed by *ftp*, *http*, or diskettes), they have a known, consistent topology and scale, and their attribute data in NED has consistent quality categories.

The standards chosen affect the infrastructure's constituency (who can take part in it) and its performance (what they can do with it). The chief goal in choosing among various standards is to find a balance between the breadth of the audience served and the depth of functions delivered. It's relatively easy to make the standards choices needed to provide high-level information functions to a well-known, homogeneous, stable audience; or simple functions to a wide, uncertain, changing audience. Choosing standards that balance reaching a large audience with providing useful functions is a difficult strategic choice for each infrastructure. The choices will (must) also evolve with changing technology, as new standards become commonplace.

The next three sections present some examples of geographic information infrastructures, the standards choices they represent, and the infrastructures that result. This is followed by some comments on evolving a standard for a changing context.

a. National Geospatial Data Clearinghouse

Since the early 1990s, the Federal Geographic Data Committee has shepherded the development of a distributed National Spatial Data Infrastructure (NSDI), with standards applied to building a National Geospatial Data Clearinghouse (cf. Chapter 8).

The current NSDI Clearinghouse, with its focus on metadata, follows a "reference library" metaphor. Users look up the basic descriptors of an item of spatial information in the Clearinghouse: whether it exists, whether a particular item will suit their needs, and where to obtain it. They then request the dataset by telephone, e-mail, or fax from the appropriate librarian; and obtain a copy of it to read or use, often with some paper trail and an agreement about acceptable use. By standardizing metadata elements (FGDC Standard) and the query protocol (Z39.50), the Clearinghouse facilitates searching for data items by allowing a single client to search many different sources at once. Putting Clearinghouse search forms on the Web makes them accessible to a wide set of users with little or no adjustments required on users' part (though some changes were required to adapt the Z39.50 to geographical searching). By (thus far) omitting a standard for retrieving the data items themselves, the Clearinghouse maintains a hands-off policy (for now) on data access and distribution: custodians of data may store it in whatever form is most convenient, and may choose to release it for a fee, license it, etc. By not standardizing data access methods, the Clearinghouse includes many data providers in the Clearinghouse who would otherwise be hesitant or otherwise ill-equipped to release their data; but it excludes casual, occasional users (who are less willing to make a formal request or to wait for delivery). Consider the prototype reported in Chapter 8: although the orthophotos were available on CDs from MassGIS, several municipal agencies chose to use the more immediately available orthophoto service we built, rather than make a telephone call and fill out order forms—and then have to worry about processing them for use. A future Clearinghouse, with direct, standard "access paths" to information resources (and access fees or user restrictions), may allow the Clearinghouse to serve a much wider audience as a reference desk for data services and not just for static data.

b. OpenGIS

Shifting to online geographic data services is more than just retrieving entire datasets: it implies querying and otherwise using remote geographic from client software system via some query protocol. Where client and server are independently designed, scalable solutions require standards to define the interaction between machines. To this end, the Open Geodata Interoperability Specification, or OpenGIS, defines standard data models and geographic processing services using an object-oriented model. The primarily industrial OpenGIS Consortium has thus far released an Abstract Specification and is currently (in 1997) applying it to data catalogs and image data, with semantic translators to follow (OGC, 1996). This standard promises to unify different vendor-specific client and server software tools and to allow users from many different organizations and industries to share a global "data space" (Buehler and McKee, 1996). OpenGIS-style interoperability would clearly make geographic information sharing more immediate and available in a wide range of inter-organizational environments. In particular, this sort of interoperability would make Chapter 8's orthophoto prototype accessible from a wide variety of software systems instead of merely a Web browser.

The chief drawback of OpenGIS is that all its benefits are expressed in the future tense. Indeed, defining a universal set of geographic data types and services is a lengthy undertaking, which has taken several years already with only minor impact on software systems as yet. One important reason for this is the newness and complexity of geographic information systems: the generic geoprocessing services at the heart of OpenGIS are still being defined, so codifying them is difficult and may even be premature. Nonetheless, ongoing coordination between the OpenGIS consortium and NSDI efforts will be important in preparing the NSDI for a networked data services environment, for instance by providing the next level of data access for the NSDI Clearinghouse.

c. Proprietary protocols: GIS-network integration

Third, the case studies and prototype highlight an emerging need to integrate GIS with network protocols. This includes both networked data access built onto GIS software, and geographic data services built into network access such as the Web. In 1996-1997, several GIS vendors began devising a variety of client-server "layers" to build networked data access into their software. These proprietary client-server tools include the Environmental Systems Research Institute's Internet Map Server (ESRI, 1997), MapInfo's ProServer (MapInfo Corp., 1997), and Intergraph's GeoMedia (Intergraph, 1997). The query protocols underlying these systems are not standards in the sense of the NSDI or OpenGIS: they are less openly published or documented, and their design is decided and controlled by one vendor rather than through industry consensus. Their advantages are those of any turnkey system: quicker start-up, more predictable deployment, and clear support by the vendor. Traditionally, these are "closed" systems of fixed components linking client and server: they require a particular GIS server, with its special networking interface, a unique protocol, a client networking interface and so on. In exchange, however, a much greater set of query and analysis functions are supported. As long as all partners in an infrastructure use the same set of software, the proprietary approach may work: for instance, in the Pacific Northwest case in Chapter 6, ESRI's Arc/Info is well established among the infrastructure participants, so the ESRI client-server suite might be promising. The case studies do show an instance of proprietary networked GIS solutions in use, in the Great Lakes: the RAPIDS air-pollution inventory team has experimented with both MapInfo's ProServer and ESRI's ArcView Internet Map Server. The vendor-specific approach may fare well in sharing information among the RAPIDS team, which is small and has a very specific focus on air pollution information, sharpened still by the three years they spent jointly developing RAPIDS software under a joint grant from EPA. It may not be as viable a solution for the more diverse, loosely connected structure of GLIN as a whole, let alone the Gulf of Maine EDIMS. Indeed, the case studies suggest that such "bridging" between independent information systems is frequently needed for inter-organizational information sharing, given that different agencies are unlikely to converge on a single set of software or data conventions. Also, Chapter 8's orthophoto prototype shows the advantages of mixing various client and server tools to use the best component for each function. Thus, vendor-specific software suites may enable rapid, predictable installation within corporate "Intranets," and may speed the deployment of existing GIS information resources over networks. Yet the promise of proprietary systems for inter-organizational information sharing may be limited: they don't allow much diversity in data systems, and specific pieces can't be swapped in and out as needed. One shift in recent years has been a blurring of standards *vs.* proprietary solutions, whereby a vendor may publish the details of a communications protocol or application program interface.

The preceding examples consist of adding networking functions to a Geographic Information System. In contrast, several firms are tucking their GIS behind increasingly sophisticated World-Wide Web sites with a mapping component. These sites, accessible to any user with a Web browser, perform geographic services such interactive mapping and trip routing. Early pioneers in this area were the Xerox map server (*mapweb.parc.xerox.com*) and the US Census Bureau's TIGER Mapping Service (*tiger.census.gov*). Commercial examples now include *www.MapQuest.com*, by Geo-Systems; *www.MapBlast.com*, by Vicinity; and *www.MapsOnUs.com*, by Lucent Technologies; related services include client-server address-matching (*www.Geocode.com*, by Etak). Some of these services may have only a limited role as infrastructure components in their current form, given that their only mode of use is an interactive HTML forms interface, using *ad hoc* query parameters. Thus, they cannot perform batch processing or function as callable objects embedded in other software systems. But this is not far off: several sites do provide maps that can easily be embedded into other Web pages (e.g., *MapBlast*), or batch services driven by special client software (e.g., *Geocode.com*). All of these sites perform specific user services reliably, and have become quite popular among a wide cross-section of users. While not supporting much interaction with client software, nonetheless they perform well at the specific service they provide. They also provide unique prototyping opportunities: their specific services (routing, geocoding, etc.), once fully developed, could be made accessible through more standardized access-paths such as OpenGIS. The lessons learned from these sites may prove useful in building unbundled, distributed GIS services.

Both the Great Lakes and the Pacific Northwest cases now feature "home built" interactive Web-based mapping tools: StreamNet's Map Builder *(http://www.streamnet.org)*, CIESIN's Great Lakes

Map Server *(http://epaserver.ciesin.org/arc/map-home.html)*, and the EPA's Maps On Demand *(http://www.epa.gov/enviro/html/mod)*. As firms like GeoSystems, Vicinity, and others turn mapping-on-the-Web into more of a commodity, these agencies, and others like them, may be able to shift their development efforts beyond simple map interfaces and towards more focused problems of data and software interoperability between client and server machines.

d. Evolution of a standard

One important challenge, highlighted by the Pacific Northwest case study in particular, is to keep readjusting the choice of strategic standards in light of new technologies, conventions, (or other standards) becoming widespread. As Pacific Northwest states began to do a lot more work with GIS, their expectations rose sharply and they demanded standards at a much more detailed scale than the 1:250,000 Reach Files. Bonneville did anticipate this several years in advance and began building more detailed Reach Files at 1:100,000; but if they've been able to standardize on something else entirely, such as dynamically segmented events or even just coordinate pairs, the outcome might have been quite different. Generally, choosing a package of strategically chosen standards is not only difficult, context-specific, and rich in consequences, it can also be made obsolete quickly by changes in the technological context. There are at least three ways to approach this challenge.

One approach involves setting a standard that is well ahead of current technology, to avoid being obsolete at the time of implementation. (Though it's not a perfect analogy, this might be called the "Intel approach," after the leading chipmaker's strategy of funding not one, but two development teams to keep it two chip generations ahead of the market (Peters, 1992).) For instance, the Information Resources Dictionary System (IRDS) (Law, 1988), defined for data repositories in the mid-to-late 1980s, employed concepts that only became common parlance five or ten years later with the growth of data warehousing. Not surprisingly, IRDS designers at the National Institute of Standards and Technology had difficulty getting users and implementers to comprehend the standard, much less comply with it (Newman, 1992); several years later, only a few repository systems even claimed to be IRDS-compliant. The OpenGIS Specification resembles this example in its complexity and its long time horizon; however, unlike the IRDS, OpenGIS is being built by an industry consortium rather than a government laboratory: this may ease its adoption by developers, but it could also succumb to competitive forces.

Another approach to evolving a standard is to limit its use to only a small community, so as to prototype it without getting locked into one version by a large base of users. One might call this the "X9 approach," after the X Window System, whose designers had the luxury of building nine different versions of it before they released it to the public (Scheifler and Gettys, 1992). (These versions were built for only a handful of research groups: this gave X developers room to define and work out a vendor-neutral, asynchronous protocol and server, without yet having to support a wide variety of uses. When X10 was publicly released, it was widely adopted; and a prior choice to keep X freely available allowed X11 to become an industry standard, maintained by the X Consortium.) This approach might provide a tangible, well-defined alternative to the comprehensive and abstract OpenGIS standard, which would speed the development and adoption of standard interfaces for sharing geographic information.

A third approach to evolving a standard is an extensible, "layered" approach, in which several specific functional standards define how components work together in an open system. As Solomon and Rutkowski (1992) describe for the case of telecommunications, "layered" standards can accommodate new features as new needs or technologies arise (new datatypes, longer addresses, wireless media, etc.)—in contrast with dimensional standards that define all communication levels at once and rely on specific pieces of hardware or software. Two prime examples of layered standards are the Internet's TCP/IP and the Open Systems Interconnection (OSI), which define independent protocols for physical transport, network connectivity, and applications. Another example is the X Window System, mentioned above: despite user pressure to broaden X's scope, its developers kept it tightly focused on the protocol and server design, deliberately leaving "details" of user interfaces or inter-client communication to individual vendors or to other standards.

In the case-study contexts of previous chapters, this "layered" approach may hold the most promise for a set of standards that can be devised and implemented quickly, and extended as new technologies become available and new standards begin to take hold. These agencies are not information-systems researchers or data suppliers: few can afford the lengthy experimentation that led to the X standard; or commit significant staff time to defining standards far in advance of current technology. Yet it's those closest to the information and its use that are most likely to decide which elements of the data sharing process to agree on, and to define a strategic set of standards that will balance the costs of standardizing vs. the benefits of collaboration, and breadth of audience vs. depth of functions served. Indeed, as mentioned earlier, while a little standardization can go a long way, the crucial decision is not what standard to choose, but what elements of the information sharing to standardize for a particular set of participants and goals—and how to keep adjusting that choice in response to a changing context.

3. Organizational implications: towards dynamic interdependence

The case studies and prototype emphasize the value of bringing together information on different topics, from different sources, describing a given location. In so doing, the study presages a not-so-distant future in which organizations will be linked in myriad complex ways, and suggests a number of strategic choices for those who will shape organizational structures and relationships for this coming context. Nonetheless, distributed responsibilities for the collection, maintenance, and analysis of geographic information may be difficult to initiate and to sustain without some attention to the incentives, costs, and benefits of geographic information sharing to participants.

a. Inter-agency teamwork and collaboration

Chapter 7 emphasized the importance of organizational forces and choices in building inter-agency infrastructures, and showed the role of shared information in inter-organizational learning and collaboration. Chapter 8 showed some flexible ways to distribute responsibilities for information ,maintenance and analysis, and suggested the dichotomy between distributed data services and conventional organizational structures. These findings have several implications for organizational choices in an increasingly inter-related and "informated" environment: in particular, new structures are likely to emerge that may extend today's partnerships model to a more dynamic, chaotic "marketplace" of interconnected players.

i. The "shadow organization" and emerging structures

In all three cases, interview subjects were members of inter-organizational structures: Advisory Boards, Steering Committees, and the like. These organizational units were quite informal and low-profile; yet they foreshadowed an emerging set of affiliations and organizational structures based not on political geography (states or countries) but on ecological geography (watersheds or fisheries). As agencies grow to depend on shared ecological information, these unofficial structures (or "shadow organizations," in the words of one Pacific Northwest respondent) could gain prominence as organizations in their own right, with patterns of influence totally outside visible affiliations or chains of command. Witness the StreamNet Web site, *www.streamnet.org:* no separate "StreamNet organization" exists at present; rather, the organization is woven into and between current agency structures throughout the Pacific Northwest. Fixing this new placeholder domain name was a not-so-subtle announcement of a "virtual" organization, distinct from any of the region's current agencies, which may never become a separate body with its own offices, hierarchy, etc. This and other new structures would recognize and embrace interdependence as a necessity within ecosystems that cross social and political boundaries: this would be a clear departure from the values of "property, privilege, and authority" prevalent in many agencies today (Lee, 1993).

ii. From partnership to marketplace: dynamic, chaotic organizational relationships

By extrapolating from the case studies and prototype a few years into the future, it seems likely that links between organizations (partnerships) and between technologies (interfaces) will grow in number and complexity, well beyond what any one plan or architecture can account for. Indeed, as organizations begin to use interoperable information services, today's partnerships, usually fewer than a dozen entities with long-term, well-defined governance, may give way to something like Dertouzos' (1997) "information marketplace" or Hock's (1995) "chaordic" (i.e. self-organizing, nearly chaotic) systems. In this view, hundreds or even thousands of organizational entities might form ephemeral alliances as needed for myriad joint projects, with informal governance and unpredictable influences from the broader economic or political context. Moore's (1996) related "ecosystem" metaphor, discussed briefly in Chapter 7, seems particularly apt in the environmental management world of the cases. Interestingly, this ecological metaphor also appears in the context of layered standards and open systems (Solomon and Rutkowski, 1992), and fittingly describes today's rapidly-shifting networked technologies and the ever-larger set of "gateways" between them. Blundon (1997) provides an early-1997 glimpse of the "ecology" the frequent redefinitions it requires, and the difficulty of predicting future outcomes.

b. Redistributing responsibilities, costs, and benefits

An important consequence of building and using information infrastructures is a redistribution of responsibilities for collecting, maintaining, and analyzing data. This begs the question: what makes such redistribution worthwhile? And the follow-up question: when information sharing is worthwhile to a region, but not to any one agency, what mechanisms are appropriate to get the infrastructure started and to keep it going?

i. When is geographic information sharing worthwhile?

An EPA manager in the Great Lakes expressed the simplest way in which information infrastructures may seem worthwhile to their users: "I get information for free." To his disappointment, however, mere enthusiasm for what could happen in a world of shared information failed to motivate many in his agency to put their information online. In general, the costs of building an information infrastructure (particularly one for geographic information) are immediate and large: lengthy learning, giving up some privacy and control, investing in new software or hardware, and people to keep these running. In contrast, the benefits of such infrastructures are more distant and uncertain: changed decisions may not occur until the infrastructure has become part of everyone's work, which may take several years; the benefits of such changes may not accrue to any of the participants themselves; and the risk of making the wrong technical and organizational choices early on is high.

Benefits of such infrastructures to participants may be more obvious where there is a clear transboundary issue: for instance, the nature of anadromous fisheries in the Pacific Northwest made it very clear that working together and sharing information was preferable to going it alone. The Great Lakes case presents similar ecological flows between its participants; these interconnections have been "amplified" by several inter-state and international agreements over several decades. Chisholm (1989, Ch. 4) discusses this kind of interdependence in depth based on overlapping bus and subway networks. The Gulf of Maine case, however, is focused on coastal habitat and shellfish, which stay put and thus present less obvious interconnectedness. In the Gulf of Maine case, regional and federal bodies are more interested in the regional view; states and provinces have little interest in, say, trading off their habitat against that of their neighbors.

The economic view, assessing costs and benefits to participants and to their broader society, is an important consideration. Nonetheless, the decision to share information, particularly in the public sector and in the environmental arena, is also a political and social one, subject to argument, persuasion, and informal relationships. Given that decisions are made by people, not by equations, they may be affected just as strongly by a clear articulation of the need for information sharing as part of a greater "stewardship" norm (or, conversely, by a call to maintain the status quo as part of a tradition or mandate), or by interpersonal affinities (or lack thereof), as by any economic analysis.

ii. Incremental vs. radical change

Given the complexity and uncertainty of organizational and technological changes, and the significant investments required, many organizations favor an incremental approach to development rather than radical changes to tasks, processes, or organizational relations. In the cases, this incrementalism was further encouraged by the intertwined nature of technological and organizational choices. Indeed, organizational change often requires technological change, and vice versa; one rarely gets very far ahead of the other. In this view, to borrow an example from political science, "it is better to work for the possible rather than to make plans for the impossible." (Reese, 1986) Without more radical changes, resulting from a creative integration of several goals and several technologies into changed organizational structures, the incremental approach may risk building only a "scaffolding" with efficiency gains such as cost savings, rather than a structure that would allow greater effectiveness through new kinds of environmental policy decisions, or new processes for reaching these decisions. Even though more radical technological and organizational change holds promise for improved planning and policy, it's unlikely to come about easily, and may require outside pressures (and often outside resources as well). Examples of these pressures are congressional acts (this was perhaps the Northwest Power Act's most important consequence), litigation (applying the Endangered Species Act applied to spotted-owl habitat in Oregon and Washington forced the Forest Service to change its *modus operandi*; applying it to salmon may do the same for many others), or inter-agency competition (as seen in the Great Lakes between CIESIN, EPA, GLIN, and others).

iii. Maintaining infrastructures for the long term

Once these complex infrastructures are launched, and organizational change is underway, are outside pressures and resources needed to sustain the shared infrastructure? How valid is the judgment of one NED participant, "No money, no conductor"? In each of the three cases, the "hub" organization had hoped to assume less and less of the infrastructure's cost after getting it started and proving its utility to participants. Yet GLIN has continued to depend on grants from Ameritech and the Environmental Protection Agency; Bonneville funds will support most of StreamNet; and the Gulf of Maine EDIMS remains limited to minimal-cost maintenance activities. When Bonneville partially divested itself of NED, it saw each state's component of the regional infrastructure diverge (though an increasingly obsolete shared standard was partly to blame). It may be some time before these infrastructures become worthwhile to their participants, but it's a feasible scenario. As new technology allows participants to provide their information to the infrastructure more easily, and as they take a more ecosystem-based and data-centered view of their work, the costs and benefits of contributing to the infrastructure may line up much more favorably. This does depend on whether the infrastructure's choice of information and distribution channels matches that of its participants: this becomes an important design consideration for long-term sustainability.

4. Policy implications: role of government, impacts on government

Public policies on information and research will be key to facilitating flexible, interoperating infrastructure components. The prototype and especially the case studies in preceding chapters offer glimpses of likely changes in environmental policymaking as a result of shared information infrastructures: wider participation, improved consensus, and detailed scientific analysis both before and after policy decisions are made.

a. Government policy in infrastructure development

Another set of policy implications concerns what public agencies could do to foster more effective infrastructures for sharing geographic information. The next paragraphs highlight the role of government agencies in providing base information to fuel myriad applications, and in setting standards to facilitate cooperative work before the market cements a *de facto* standard.

i. Information to fuel applications; rethinking ownership

The importance of a common reference system for sharing information is unique to geographic information, and suggests an important role for government agencies in promoting geographic information infrastructures. In NSDI parlance, a Framework of digital geographic data, such as roads, hydrology, terrain, or orthophotos, will encourage the creation of other geographic data that is geographically consistent and thus increases the value of all derivative works. This government role was clearly seen in the production of the Topographically Integrated Geographic Encoding and Referencing system (TIGER), a systematic map of streets throughout the US, with topology and address ranges. Although initially devised for the 1990 Census, it has seen use and derivative works in all sectors of public and private management, and commercial mapping products. Government could play a key role in coordinating mapping efforts (both private and public) by location, so as to produce the Framework's base data layers, and to ensure that these layers are made as widely available as possible. One key mechanism for this would be seed money and cost-sharing. Seed money in particular would assist with developing initial Framework data in an area; once some "critical mass" of Framework data became available, other data providers in the area would voluntarily choose to follow Framework specifications themselves because of the added value of combining their information with that already available. Cost sharing would be one mechanism for ensuring that data are freely distributed once developed.

As information networking and data services become more widespread, producers of information will need to find new ways to define their ownership of information. This is especially true for geographic information, which is being created in large volumes mostly by public agencies. Based on experiences in the United States, Lopez (1995) groups current forms of public information dissemination into three groups: cost recovery, private distributors, and open access. He argues that only open networked access satisfies two peculiar requirements of government information: it encourages demand, and complies with the public's "right to know." In this view, networked access is key to open access in that it limits dissemination costs and makes government information immediately available to the public. Networked open-access information may require many government agencies to rethink their views on ownership of the information they've produced, including copyrights, licensing, usage, and redistribution. A move from information products to information services may require further redefinitions.

ii. Setting de jure *standards*

Standards are another important role for public agencies in fostering geographic information infrastructures. If a *de jure* standard, set by an open, public standards process, can be defined before a *de facto* standard becomes established ("gelled") through market forces, then easy interchange of data, or interoperability among services, becomes possible without giving up the competition among vendors that keeps quality high. Although standards-setting by industry consortia has had successful results as well (e.g. the X Window System), the public standards process is more open to input from a broad set of potential users. (Melton, 1996) This is an important role to be played by the public sector in fostering inter-organizational information infrastructures.

b. Infrastructures in government policy

The first policy implication of this research is its potential to improve the nature and process of environmental policy making. In particular, making detailed information more widely available promises to broaden participation in decisions and make it more conducive to mutual understanding and consensus.

i. Broader participation and more effective consensus

One first-order impact of any information infrastructure would be the chance to involve many otherwise excluded participants in the decision-making process, by providing them with detailed information, opportunities to contact key decision makers, and forums for deliberating different viewpoints. In addition to greater accessibility, current networked communication tools (electronic mailing lists, newsgroups, online chat meetings, and the like) would allow larger numbers of participants (hundreds or thousands) to take part in debates, discussions, and joint decisions. This impact was seen to varying degrees in all three cases; and was an important goal for the Environmental Protection Agency in particular.

But geographic information infrastructures would allow for much more than just efficient verbal debates. First, as proposed in Chapter 7, maps are promising "boundary objects" (Boland and Tenkasi, 1995) for articulating and reconciling different world-views. Furthermore, thanks to networked data services of the kind suggested in Chapter 8, participants might interact with dynamic models to sketch the outcomes of various assumptions or scenarios. This would also offer a natural way to break up complex physical phenomena or lines of reasoning into more easily understandable pieces. Thus, adding online interactive models to map displays may prove to be powerful facilitators of consensus in complex decisions—thus improving both the process and the outcome of policies.

ii. Improved scientific analysis before and after policies are made

A second-order benefit of geographic information-sharing infrastructures is particularly relevant to environmental policy and planning, although it would apply to any complex science-intensive decision-making. Science-intensive policy decisions depend on obtaining and interpreting trusted, high-quality, up-to-date information. Traditionally, this would be done by copying all relevant data sets to a single location, or by replicating fragments of the data sets onto myriad individual machines. In either case, the data may be misused by generalists, and is unlikely to undergo regular updates. Most commonly, however, detailed "context" data is just too difficult to obtain, and so simpler models are used, which make simplifying assumptions about existing conditions. Information-sharing infrastructures would allow detailed information to be left in the care of those who collect it, maintain it, and know it best. This would maintain its accuracy and proper use; while allowing users to obtain the information directly from the source when needed. Detailed information for modeling would be helpful not only prior to policy decision, but afterwards as well, so as to evaluate the effect of policies and thus use policy measures as a means to learn about an ecosystem's complex behavior. This is a key emphasis of so-called *adaptive management*, as put forth by Lee (1993), an important goal expressed particularly in the Pacific Northwest given that region's long struggle to manage a still poorly understood ecosystem.

5. Hypotheses for further research

Many of the implications sketched in this chapter rest on predictions about the future; they leave several questions unanswered about technology, organizations and policy related to geographic information infrastructures. Thus, this study suggests a number of fruitful areas for future research. The following is a list of some of the most crucial hypotheses to be tested, culled from the likely scenarios sketched in various parts of this chapter and the previous ones.

1. Because the organizational and technological sides of a geographic information infrastructure are interdependent, parallel, incremental change in both seems to be the most likely pattern. Is this necessarily the case? What opportunities might there be for "leapfrogging" over several development stages, to acquire the needed complexity much more quickly?

2. Organizations that can undergo "deeper" changes in structure, in favor of interdependence with other agencies based on geographic information, stand to benefit more from geographic information infrastructures. What are these deep changes, and the impacts to be expected?

3. Geographic information infrastructures are leading from today's partnerships to more ephemeral relationships among large groups of organizations that share a particular region or location. Are there emerging examples of this already? What kinds of norms, rules, incentives, and standards would keep these infrastructures together? How dynamic and chaotic can such a geographic information infrastructure get, while still allowing its participants to benefit from each other's geographic information?

4. Given that a comprehensive functional standard such as OpenGIS is such a difficult undertaking, partial functional specifications would seem important in the short- to medium-term in facilitating the sharing of geographic information between independent systems. What are the most crucial elements of such a partial specification?

5. What long-term institutional structures are appropriate for geographic information infrastructures? Can they ever be self-sustaining, or will they always require significant outside support?

These and other questions would enlighten the future development of geographic information sharing infrastructures, and their role in devising improved public policy and planning based on the best available information.

Bibliography

Adhikari, Richard, 1997. Envirofacts Assists EPA Programs. *Information Week*, March 17, 1997, p. 95-96. Article online at URL *<http://techweb.cmp.com/iw/622/22mtweb.htm>.*

Alexander, Ernest R., *How organizations act together: interorganizational coordination in theory and practice.* Australia, United States: Gordon and Breach.

Allen, Stan, O'Connor, Dick, Roger, Phillip B., Beamesderfer, Ray, and Rowe, Mike, *Columbia River Coordinated Information System: FY 1994 Annual Report.* Unpublished; available from Pacific State Marine Fisheries Commission, Gladstone, OR.

Attewell, Paul, and Rule, James, 1984. Computing in organizations: what we know and what we don't know. *Communications of the ACM,* Vol. 27, No. 12, pp. 1184-1192.

August, Peter V., 1991. Use vs. abuse of GIS data. *URISA Journal,* Vol. 3, No. 2, pp. 99-101.

Baker, Tony, and Broadhead, John, 1992, Integrated access to land-related data from heterogeneous sources. In Salling Mark J. (ed.), *Proceedings of the 1992 Annual Conference of the Urban and Regional Information Systems Association (URISA).*

Barley, Stephen R., 1986. Technology as an occasion for structuring: evidence from observations of CT scanners and the social order of radiology departments, *Administrative Science Quarterly,* Vol. 31, 1986, pp. 78-108.

Batini, C, Lenzerini M., and Navathe, S., 1986, A comparative analysis of methods for database schema integration, *Association for Computing Machinery (ACM) Computing Surveys,* Vol. 18, No. 4, pp. 323-364.

Beaulac, Michael N., Businski, Shirley, and Forstat, David, 1994. The evolution of Michigan's geospatial data infrastructure, *Journal of the Urban and Regional Information Systems Association,* Vol. 6, No. 1, Spring 1994, pp. 63-68.

Berners-Lee, T. J., Cailliau, R., and Groff, J.-F., 1992. The world-wide web. *Computer networks and ISDN systems,* Vol. 25, No. 4-5, p. 454.

Blumenthal, Marjory S., 1997. Unpredictable certainty: the Internet and the information infrastructure, *IEEE Computer,* January 1997, pp. 50-56.

Blundon, William, 1997. How Microsoft is broadening Java's scope. *JavaWorld,* Vol. 2, No. 2.

Boland, Richard J., and Tenkasi, Ramkrishnan V., 1995. Perspective making and perspective taking in communities of knowing. *Organization Science* Vol. 6, No. 4, pp. 350-372.

Bonneville Power Administration (BPA), 1985. *Geographic Information System Task Force: Report and Recommendations.* Division of Power Resources Planning, Office of Power and Resources Management. Portland, Bonneville Power Administration (unpublished).

Bonneville Power Administration (BPA), 1993. *Resource Programs Final Environmental Impact Statement. Volume 1, Environmental Analysis.* Section 2.2.7, "Protected Areas," p. 28. Portland, OR: Bonneville Power Administration.

Bonneville Power Administration (BPA), 1996. *StreamNet: Report on the status of salmon and steelhead in the Columbia River basin, 1995.* Portland, OR: Bonneville Power Administration.

Brodie, Michael L., 1993. The promise of distributed computing and the challenges of legacy information systems. In Hsiao, David K., Neuhold, Erich J., and Sacks-Davis, Ron (eds.), *Interoperable Database Systems (DS-5),* pp. 1-31. North-Holland: Elsevier.

Brown, Wendell S., Garrison, Karen M., and MacBeth, Mark, 1994. *Prototype Environmental Data and Information Management Systems (EDIMS). <http://gulfofmaine.unh.edu/edims/edims_description.html>*

Buehler, Kurt, and McKee, Lance (eds.), 1996. The OpenGIS Guide: Introduction to interoperable geoprocessing, Part I of the Open Geodata Interoperability Specification (OGIS). Available from the Open GIS Consortium, Inc., Wayland, Massachusetts. URL *<http://www.-opengis.org/public/guide.html>*

Carlson, W. Bernard, 1992. Artifacts and frames of meaning: Thomas A. Edison, his managers, and the cultural construction of motion pictures. In Bijker, Wiebe E., and Law, John (eds.), *Shaping Technology/Building Society: Studies in Sociotechnical Change*, pp. 175-198. Cambridge, MA: MIT Press.

Carter, James R., 1992. "Perspectives on sharing data in geographic information systems," *Photogrammetric Engineering and Remote Sensing*, Vol. 58, No. 11, November 1992, pp. 1557-1560.

Chisholm, Donald, 1989. *Coordination without hierarchy: informal structures in multiorganizational systems.* Berkeley, CA: University of California Press.

Chorley, Roger, 1988. Some reflections on the handling of geographic information. *International Journal of Geographic Information Systems*, Vol. 2, No. 1, pp. 3-9.

Christian, Eliot J.,1996. GILS: What is it? Where's it going? *D-Lib Magazine*, URL *<http://www.-dlib.org/dlib/december96/12christian.html>*.

Cobb, T. E., and Seals, J. E., 1996. The Gulf of Mexico Information Network (GIN). *Proceedings of Eco-Informa '96: global networks for environmental information.* Ann Arbor, MI: Environmental Research Institute of Michigan. Cf. URL *<http://pelican.gmpo.gov>*.

Cooke, Donald F., 1995, "Sharing street centerline spatial databases," Chapter 22 of Onsrud, Harlan J., and Rushton, Gerard (eds.), *Sharing Geographic Information.* New Brunswick, NJ: Center for Urban Policy Research, Rutgers University.

Consortium for International Earth Science Information Network (CIESIN), 1996. *What is CIESIN?* URL *<http://www.ciesin.org/about/CIESIN-Page.html>*.

Consortium for International Earth Science Information Network (CIESIN), 1996b. *GLREIS home page.* URL *<http://epaserver.ciesin.org>*.

Craig, William J., 1995. Why we can't share data: institutional inertia. Chapter 6 of Onsrud, Harlan J., and Rushton, Gerard(eds.), *Sharing Geographic Information.* New Brunswick, NJ: Center for Urban Policy Research, Rutgers University.

Croswell, Peter L., 1989. Facing reality in GIS implementation: lessons learned and obstacles to be overcome. In *Proceedings of the 1989 Annual Conference of the Urban and Regional Information Systems Association (URISA)*, edited by Salling, Mark J. and Gayk, Williams F., Vol. IV (Washington, DC: URISA), pp. 15-35.

Dertouzos, Michael L., 1997. Creating the people's computer. *Technology Review*, Vol. 100, No. 3, April 1997, pp. 20-28.

Dertouzos, Michael L., 1997b. *What will be: how the new world of information will change our lives.* San Francisco: HarperEdge.

DeSanctis, Geraldine, and Poole, Marshall Scott, 1994. Capturing the complexity in advanced technology use: adaptive structuration theory. *Organization Science*, Vol. 5, No. 2, May 1994.

Dietrich, William, 1995, *Northwest Passage: the great Columbia River.* New York: Simon & Schuster.

Dueker, Kenneth J., and Vrana, Ric, Systems integration: a reson and a means for data sharing, Chapter 9 of Onsrud, Harlan J., and Rushton, Gerard (eds.), *Sharing Geographic Information.* New Brunswick, NJ: Center for Urban Policy Research, Rutgers University.

Egenhofer, Max J. 1992. "Why not SQL!", *International Journal of Geographic Information Systems,* Vol. 6, No. 2, pp. 71-85.

Emtage, Alan, and Deutsch, Peter, 1992. *Archie*: an electronic directory service for the Internet. In *Proceedings of the Winter 1992 USENIX Conference,* pp. 93-110. Berkeley, California: The USENIX Association.

Environmental Protection Agency (EPA), 1996. *Region 5 automated spatial data library maintained by Office of Information Services.* URL *<http://www.epa.gov/reg5ogis/catalog.htm>.*

Environmental Protection Agency (EPA), 1997. *Overview of the (legacy) STORET system.* URL *<http://www.epa.gov/OWOW/STORET/overview.html>.*

Environmental Protection Agency (EPA), 1997b. *Welcome To Envirofacts.* URL *<http://www.epa.gov/enviro/>.*

Environmental Protection Agency (EPA), 1997c. *Site based mapping application (National SITE-INFO).* URL *<http://www.epa.gov/region10/www/gisapps/natsite.html>.*

Environmental Protection Agency Great Lakes National Program Office (EPA-GLNPO), 1995. *The Great Lakes: An environmental atlas and resource book.* Chicago, Ill.: EPA Great Lakes National Program Office. Toronto, Ont.: Government of Canada. Electronic version available at URL *<http://www.epa.gov/glnpo/atlas/intro.html>.*

Environmental Systems Research Institute (ESRI), 1997. *Internet mapping solutions.* URL *<http://maps.esri.com/ESRI/esri.htm>.*

Epstein, Earl F., 1995, Control of public information. Chapter 17 of Onsrud, Harlan J., and Rushton, Gerard (eds.), *Sharing Geographic Information.* New Brunswick, NJ: Center for Urban Policy Research, Rutgers University.

Evans, John D., Ferreira, Joseph Jr., and Thompson, Philip R., 1992, A visual interface to spatial databases based on spatial metadata. In *Proceedings of the 5th International Symposium on Spatial Data Handling,* Bresnahan, P., Corwin, E., and Cowen, D. (eds.), Vol. 1, pp. 282-293. Charleston, SC: International Geographical Union (IGU) and University of South Carolina.

Evans, John D., and Ferreira, Joseph Jr., 1995. "Sharing spatial information in an imperfect world: interactions between technical and organizational issues," Chapter 27 of Onsrud, Harlan J., and Rushton, Gerard (eds.), *Sharing Geographic Information.* New Brunswick, NJ: Center for Urban Policy Research, Rutgers University.

Executive Office of the President (EOP), 1993. *The President's Forest Plan: breaking gridlock and moving forward. <http://www.whitehouse.gov/WH/EOP/OP/html/forest.html>*

Executive Office of the President (EOP), 1994. Coordinating geographic data acquisition and access: the National Spatial Data Infrastructure (Presidential Executive Order No. 12906). *Federal Register,* Vol. 59, No. 71 (April 13, 1994), pp. 17671-17674.

Federal Geographic Data Committee (FGDC), 1994. *Content standards for digital geospatial metadata.* Washington, DC: Federal Geographic Data Committee.

Federal Geographic Data Committee (FGDC), 1995. *Development of a National Geospatial Data Framework.* Washington, DC: Federal Geographic Data Committee.

Firns, Peter G., 1992, On the notion of a spatial entity in the context of data modeling for spatial information systems. *Proceedings, Fourth Colloquium of the Spatial Information Research Center, Dunedin, New Zealand, 18th-20th May, 1992.*

Fisher, Bruce J., 1993. Methodology used to produce an encoded 1:100,000 scale River Reach data layer for the Pacific Northwest, in *Pacific Northwest River Reach Files: reach documentation.* Open file report. Unpublished; available from State Service Center for Geographic Information Systems, Oregon Dept. of Administrative Services, Salem, OR.

Forsberg, Brent O., 1992. *Oregon Rivers Information System Operation Manual.* Unpublished; available from Oregon Dept. of Fish and Wildlife, Portland, OR.

Frank, Andrew U., 1992. Acquiring a digital basemap: a theoretical investigation into a form of sharing data. *Journal of the Urban and Regional Information Systems Association,* Vol. 4, No. 1, Spring 1992, pp. 10-23.

Frank, Steven, 1994. The National Spatial Data Infrastructure: designing navigational strategies, *Journal of the Urban and Regional Information Systems Association,* Vol. 6, No. 1, pp. 37-55.

Frew, James, Freeston, Michael, Kemp, Randall B., Simpson, Jason, Smith, Terence, Wells, Alex, and Zheng, Qi, 1996. The Alexandria Digital Library Testbed, *D-Lib Magazine,* URL *<http://www.dlib.org/dlib/july96/alexandria/07frew.html>.*

Fulghum, Robert, 1989. *All I really need to know I learned in kindergarten,* p. 6. New York: Villard Books.

Ganter, John H., 1995, The Open Geodata Interoperability Specification (OGIS) as a technology for geospatial transportation computing. In *Proceedings of the 8th Geographic Information Systems for Transportation (GIS-T) Symposium, Sparks, Nev.,* pp. 535-553. Washington: American Association of State Highway and Transportation Officials.

Giddens, Anthony, 1984. *The Constitution of Society.* Berkeley, CA: University of California Press.

Goodchild, Michael F., 1992. Geographic Information Science. *International Journal of Geographic Information Systems,* Vol. 6, No. 1, pp. 31-45.

Glaser, Barney G., and Strauss, Anselm L., 1967. *The discovery of grounded theory.* Chicago: Aldine Publishing Co.

Grandori, Anna, and Soda, Giuseppe, 1995. Inter-firm networks: antecedents, mechanisms, and forms. *Organization Studies,* Vol. 16, No. 2, pp. 183-214.

Granovetter, Mark S., 1972. The strength of weak ties. *American Journal of Sociology,* Vol. 78, No. 6, pp. 1360-1380.

Gray, Barbara, 1985. Conditions facilitating interorganizational collaboration. *Human Relations,* Vo. 38, No. 10, pp. 911-936.

Harrigan, Kathryn Rudie, and Newman, William H., 1990. Bases of Interorganizational cooperation: propensity, power, persistence. *Journal of Management Studies,* Vol. 27, No. 4, pp. 417-434.

Hock, Dee W., 1995. The Chaordic Organization: Out of Control and Into Order, *World Business Academy Perspectives,* Vol. 9, No.1, pp. 5-21.

Hornadle, George, and Cooper, Lauren, 1989. Beyond coordination and control: an inter-organizational approach to structural adjustment, service delivery, and natural resource management, *World Development,* Vol. 17, No. 10, pp. 1531-1541.

Houghton-Mifflin, 1992. *The American Heritage Dictionary of the English Language, Third Edition.*

Hsu, Cheng, Bouziane, M'hamed, Rattner, Laurie, and Yee, Lester, 1991. Information resources management in heterogeneous, distributed environments: a metadatabase approach. *Institute of Electrical and Electronics Engineers (IEEE) Transactions on Software Engineering,* Vol. 17, No. 6, pp. 604-624.

Idaho Department of Fish and Game (IDFG), 1994. *Idaho Rivers Information System: user documentation.* Unpublished; available from Idaho Dept. of Fish and Game, Boise, ID.

Intergraph Corporation, 1997. *GeoMedia Web page.* URL <*http://www.intergraph.com/iss/geomedia*>.

International Joint Commission (IJC), 1995. *The Great Lakes: a vital resource worth protecting.* Brochure available from the International Joint Commission, Windsor, Ont.

Jones, Mark R., 1992. Unveiling repository technology, *Database Programming and Design*, Vol. 5, No. 4, pp. 28-35.

Kahle, Brewster, 1991. An information system for corporate users: Wide Area Information Servers, *Online Magazine*, Vol. 15, No. 5, pp. 56-62.

King, J. L., and Kraemer, K., 1980. *Cost as a social impact of telecommunications and other information technologies.* Available from Public Policy and Research Organization, Irvine, CA

Knudsen, K. Lea, Jackson, Steve, Lovgren, Terry, Hunter, Mark, and McDonald, Dennis, 1992. *Washington Rivers Information System, resident and anadromous fish data: 1:100,000 scale update.* Unpublished; available from Washington Dept. of Fish and Wildlife, Olympia, WA.

Law, Margaret Henderson, 1988. *Guide to Information Resource Dictionary System applications: general concepts and strategic systems planning.* National Bureau of Standards (now National Institute of Standards and Technology) Special Publication No. 500-152. Gaithersburg, MD: NIST.

Lee, Kai S., 1993. *Compass and Gyroscope: integrating science and politics for the environment.* Washington: Island Press.

Litwin, Witold, Mark, Leo, and Roussopoulos, Nick, 1990, Interoperability of multiple autonomous databases, *Association for Computing machinery (ACM) Computing Surveys*, Vol. 22, No. 3, pp. 267-293.

Longman Group UK, 1994. *The Original Roget's Thesaurus of English Words and Phrases* (Americanized Version). London: Longman Group.

Lopez, Xavier R., From gravel to diamonds: the National Spatial Data Infrastructure at a crossroads. In *Proceedings of the 1995 Annual Conference of the Urban and Regional Information Systems Association (URISA)*, pp. 611-625.

Mackay, D. Scott, and Robinson, Vincent B., 1992. *Towards a heterogeneous information systems approach to geographic data interchange.* Institute for Land Information Management Discussion Paper 92/1 (Mississauga, Ontario: University of Toronto).

Mapinfo Corporation, 1997. *MapInfo ProServer.* URL <*http://www.mapinfo.com/events/prosrv/-proserver.html*>.

Markus, M. Lynne, and Robey, Daniel, 1988. Information technology and organizational change: causal structure in theory and research, *Management Science*, Vol. 34, No. 5, pp. 583-598.

Melton, Jim, 1996. Why standardize? A response to the arguments against standardization. *Database programming and design*, April 1996, pp. 60-63.

Menke, William, Friberg, Paul, Lerner-Lam, Arthur, Simpson, David, Bookbinder, Robert, and Kerner, Garry, 1991. A voluntary, public method for sharing earth science data over Internet using the Lamont view-server system. *Eos, American Geophysical Union Transactions*, July 1991, pp. 409-414.

Merriam-Webster, Inc., 1990. *Webster's Ninth New Collegiate Dictionary.* Springfield, MA: Merriam-Webster.

Misa, Thomas J., 1994. Retrieving Sociotechnical change from technological determinism, in Smith, Merrit Roe, and Marx, Leo, *Does technology drive history? The dilemma of technological determinsim*, pp. 115-141. Cambridge, MA: MIT Press.

Mizruchi, Mark S., and Galaskiewicz, Joseph, 1993. Networks of interorganizational relations. *Sociological Methods and Research*, Vol. 22, No. 1, pp. 46-70.

Montana Department of Fish, Wildlife, and Parks (DFWP), 1993?. *The Montana Rivers Information System.* Brochure available from Montana DFWP, Kalispell, MT.

Montana Department of Fish, Wildlife, and Parks (DFWP), 1994. *Montana Rivers Information System Reporting User's Manual.* Available from Montana DFWP, Kalispell, MT.

Montana Natural Resource Information System (NRIS), 1996. *Montana Rivers Information System (MRIS).* URL <*http://nris.mt.gov/wis/mris1.html*>.

Moore, James F., 1996, *The Death of Competition: leadership and strategy in the age of business ecosystems.* New York: HarperCollins.

Morehouse, Scott, 1990, The role of semantics in geographic data modeling. *Proceedings of the 4th International Symposium on Spatial Data Handling, Zurich, Switzerland,* pp. 689-698.

Narayan, Rom, 1988. *Data Dictionary: Implementation, Use, and Maintenance.* Englewood Cliffs, NJ: Prentice-Hall.

National Institute of Standards and Technology (NIST), 1992. *Spatial Data Transfer Standard (SDTS). Federal Information Processing Standards (FIPS) Publication 173.* Washington, DC: US Government Printing Office.

National Research Council, 1983. *Procedures and Standards for a Multipurpose Cadastre.* Washington, D.C.: National Academy Press.

National Research Council, 1993. *Toward a coordinated spatial information infrastructure for the nation.* Mapping Science Committee, Board on Earth Sciences and Resources, Commission on Geosciences, Environment, and Resources. Washington, DC: National Academy Press.

National Research Council, 1994. *Promoting the National Spatial Data Infrastructure through partnerships.* Mapping Science Committee, Board on Earth Sciences and Resources, Commission on Geosciences, Environment, and Resources. Washington, DC: National Academy Press.

National Research Council, 1995. *Upstream: Salmon and Society in the Pacific Northwest* (pre-publication copy). Committee on Protection and Management of Pacific Northwest Anadromous Salmonids, Board on Environmental Studies and Toxicology, Commission on Life Sciences. Washington, DC: National Academy Press.

National Telecommunications and Infrastructure Administration, 1996. *Notice of Solicitation of Grant Applications, Telecommunications and Information Infrastructure Assistance Program.* URL <*http://www.ntia.doc.gov/otiahome/tiiap/96notice.html*>

Nebert, Douglas, 1995. *What does it mean to be a Clearinghouse node?* Draft position paper for FGDC Clearinghouse working group, December 1995.

Newman, Ian A., 1992, Data dictionaries, Information Resource Dictionary Systems, and metadatabases. In Medyckyj-Scott, David J., Newman, Ian A., Ruggles, Clive L. N., and Walker, David R. F. (eds.), 1992, *Metadata in the Geosciences,* pp. 69-83. Loughborough, UK: Group D Publications Ltd.

Northwest Power Planning Council (NWPPC), 1988. *Protected Areas Amendment and Response to Comments,* September 14, 1988, Publication 88-22. Portland, OR: Northwest Power Planning Council.

Nyerges, Timothy L., 1989. Schema integration analysis for the development of GIS databases, *International Journal of Geographical Information Systems,* Vol. 3, No. 2, pp. 153-183.

Nyerges, Timothy L., 1991, Geographic information abstractions: conceptual clarity for geographic modeling. *Environment and Planning A,* Vol. 23, pp. 1483-1499.

Obermeyer, Nancy J., 1995. Reducing inter-organizational conflict to facilitate sharing geographic information. Chapter 8 of Onsrud, Harlan J., and Rushton, Gerard (eds.), *Sharing Geographic Information.* New Brunswick, NJ: Center for Urban Policy Research, Rutgers University.

Office of Management and Budget (OMB), 1990. *Coordination of surveying, mapping, and related spatial data activities*, Circular A-16. Washington, DC: Office of Management and Budget.

Open GIS Consortium (OGC), 1996. *OGC Technical Committee Request Plan & Schedule, Revision 1.* OGIS Project Document No. 96-011R1. URL <*http://www.opengis.org/public/request.html*>.

Orlikowski, Wanda J., 1992. The duality of technology: rethinking the concept of technology in the context of organizations, *Organization Science*, Vol. 3, No. 3, pp. 398-427.

Pacific States Marine Fisheries Commission (PSMFC), 1997. *PSMFC Projects.* URL <*http://www-.psmfc.org/projects.html*>

Pascoe, R. T., and Penny, J. P., 1990. Construction of interfaces for the exchange of geographic data, *International Journal of Geographic Information Systems,* Vol. 4, No. 2, pp. 147-156.

Peckham, Joan, and Maryanski, Fred, 1988, Semantic data models, *Association for Computing machinery (ACM) Computing Surveys*, Vol. 20, No. 3, pp. 153-189.

Peters, Tom, 1992. *Liberation Management: necessary disorganization for the nanosecond nineties.* New York: Knopf.

Pinch, Trevor J., and Bijker, Wiebe E., 1984. The social construction of facts and artefacts: or how the sociology of science and the sociology of technology might help each other. In Bijker, W., Hughes, T., and Pinch, T. (eds.), *Social Studies of Science*, Vol. 14, pp. 399-441.

Pinto, Jeffrey K., and Onsrud, Harlan J., 1995. Facilitators of organizational information sharing: a research framework. Chapter 3 of Onsrud, Harlan J., and Rushton, Gerard (eds.), *Sharing Geographic Information.* New Brunswick, NJ: Center for Urban Policy Research, Rutgers University.

Pinto, Mary Beth, Pinto, Jeffrey K., and Prescott, John E., 1993. Antecedents and consequences of project team cross-functional cooperation, *Management Science*, Vol. 39, No. 10, pp. 1281-1297.

Ralston, Patrick R., 1995. A legacy of cooperation. In *ADVISOR* (Great Lakes Commission newsletter), Vol. 8, No. 1, p. 3.

Ratza, Carol A., 1996. The Great Lakes Information Network: the region's Internet information service, *Toxicology and Industrial Health*, Vol. 12, Nos.3/4, pp. 557-561.

Reese, Thomas J., 1986. Miracles or incrementalism, *National Tax Association Journal*, No. 39.

Rhind, David W., 1992. Data access, charging, and copyright and their implications for geographical information systems. *International Journal of Geographical Information Systems,* Vol. 6, No. 1, pp. 13-30.

Robinson, Vincent B., and Sani, Anthony P., 1993. Modeling geographic information re-sources for airport technical data management using the Information Resources Dictionary System standard, *Computers, Environment, and Urban Systems,* Vol. 17, pp. 111-127.

Scheifler, Robert W., and Gettys, James, 1992. *X Window System: the complete reference to Xlib, X Protocol, ICCCM, XLFD* (3rd Edition), pp. 8-15. Burlington, MA: Digital Press.

Schermerhorn, John R., 1975. Determinants of interorganizational cooperation. *Academy of Management Journal*, Vol. 18, No. 4, pp. 846-856.

Schwartz, Michael F., Emtage, Alan, Kahle, Brewster, and Neuman, B. Clifford, 1992. A comparison of Internet resource discovery approaches. *Computing Systems: the journal of the USENIX Association,* Vol. 5, No. 4, pp. 461-493.

Sen, A. and Kerschberg, L., 1987. Enterprise modeling for database specification and design, *Data and Knowledge Engineering*, Vol. 2, 31-58.

Sheth, Amit P., and Larson, James A., 1990. Federated database systems for managing distributed, heterogeneous, and autonomous databases, *Association for Computing Machinery (ACM) Computing Surveys*, Vol. 22, No. 3, pp. 183-236.

Siegel, Jon, 1996, *CORBA fundamentals and programming*. New York: Wiley.

Siegel, Michael D., and Madnick, Stuart E., 1991, Context interchange: sharing the meaning of data. *ACM SIGMOD Record* (Association for Computing Machinery: Special Interest Group on the Management of Data), Vol. 20, No. 4, pp. 77-78.

Simon, Herbert A., 1981. *The sciences of the artificial.* Cambridge, Mass.: MIT Press.

Solomon, Richard J., and Rutkowski, Anthony M., 1992. Standards-making for IT: old vs. new models. Presented at the Conference on the Economic Dimension of Standards: Users and Governments in IT Standardization, Tokyo, 18 Nov. 1992.

Spackman, J. W. C., 1990. The networked organisation. *British Telecommunications Engineering*, Vol. 9, April 1990.

Sugarman, Barry, 1996. *Notes towards a closer collaboration between organization theory, learning organizations, and organizational learning in the search for a new paradigm.* Working Paper, MIT Center for Organizational Learning, Cambridge, Mass. URL <*http://learning.mit.edu/res/wp/Sugarman-.html*>

Templeton, Marjorie, Brill, David, Dao, Son K., Lund, Eric, Ward, Patricia, Chen, Arbee L. P., and MacGregor, Robert, 1987. Mermaid: a front-end to distributed heterogeneous databases. *Proceedings of the Institute of Electrical and Electronics Engineers (IEEE)*, Vol. 75, No. 5, pp. 695-708.

Thomas, Robert J., 1993. *What Machines Can't Do: Politics and Technology in the Industrial Enterprise*, p. 239. Berkeley, CA: University of California Press.

Tosta, Nancy, 1992. Data sharing. *Geo Info Systems*, July/August 1992, pp. 32-34.

Tosta, Nancy, 1994. Continuing evolution of the National Spatial data Infrastructure. In *Proceedings, GIS/LIS '94*. Bethesda: ACSM-ASPRS-AAG-URISA-AM/FM. Vol. 1, pp. 768-776.

Tosta, Nancy, and Domaratz, Michael, 1997. The US National Spatial Data Infrastructure. Chapter 2 in Craglia, Massimo, and Couclelis, Helen (eds.), *Geographic Information Research: Bridging the Atlantic.* London: Taylor & Francis.

Turner, Barry A., 1983. The use of grounded theory for the qualitative analysis of organizational behavior. *Journal of Management Studies*, Vol. 20, No. 3, July 1983, pp. 333-348.

U.S. Congress, 1980. *Pacific Northwest Electric Power Planning and Conservation Act*, 16 U.S. Code Chapter 12H (1988 Ed.) Act of Dec. 5, 1980, 94 Stat. 2697. Public Law No. 96-501, S. 885.

Vanzwol, Cathleen, 1995. Washington Surface Water Identification System. In *Proceedings of the 15th Annual ESRI User Conference.* Redlands, CA: Environmental Systems Research Institute. Abstract available at URL <*http://www.esri.com/base/common/userconf/proc95/to350/p346.html*>

Wall, Larry, Schwartz, Randal L., and Christiansen, Tom, 1996. *Programming Perl*, 2nd Ed. Sebastopol, CA: O'Reilly & Associates.

Walker, David R. F., Newman, Ian A., Medyckyj-Scott, David J., and Ruggles, Clive R. N., 1992. A system for identifying datasets for GIS users. *International Journal of Geographical Information Systems*, Vol. 6, No. 6, pp. 511-527.

Wang, Y. Richard, and Madnick, Stuart E., 1989. Facilitating connectivity in composite information systems. *Data Base*, Fall 1989, pp. 38-46.

Weick, Karl E., Technology as equivoque: sensemaking in new technologies. In Goodman, P., Sproul, L., et al., *Technology and Organizations,* pp. 1-44. San Francisco, CA: Jossey-Bass.

Worboys, Michael F., and Deen, S. M., 1991. Semantic heterogeneity in distributed geographic databases. *ACM SIGMOD Record* (Association for Computing Machinery: Special Interest Group on the Management of Data), Vol. 20, No. 4, pp. 30-34.

Yaffee, Steven L., 1994. *The wisdom of the spotted owl: policy lessons for a new century.* Washington, DC: Island Press.

Yates, H. J., 1990. Information resource management, *British Telecommunications Engineering,* Vol. 9, pp. 22-27.