Implementation of *The National Map* Road Database with Considerations for Integration and Organizational Constraints

Project Report
USGS Mid-Continent Mapping Center, Rolla, Missouri
August 2003

Bryan Weaver
Graduate Student
Department of Geography
University of Georgia
bweaver@uga.edu

This report was prepared under a grant from the U.S. Geological Survey to the University of Georgia.
Table of Contents

Executive Summary .................................................................................................................................. 2
1. Introduction ........................................................................................................................................ 3
   Background .......................................................................................................................................... 3
   USGS and The National Map .................................................................................................................. 4
   Research project focus ........................................................................................................................ 5
2. Methodology ....................................................................................................................................... 6
   The National Map descriptive documentation review ............................................................................ 6
   Interaction with The National Map personnel ....................................................................................... 6
   Peer review literature referencing ........................................................................................................ 7
   Data evaluation .................................................................................................................................... 7
3. Findings ........................................................................................................................................... 7
   3.1. Status of The National Map project ............................................................................................... 7
   3.2. Recommended project changes ........................................................................................................ 8
       Project recognition and consolidation .................................................................................................. 9
       Project leadership ............................................................................................................................. 10
       Project Requirements ....................................................................................................................... 10
       Project implementation .................................................................................................................... 11
       Taxonomy of systems integration: .................................................................................................... 12
   3.3. Integrating The National Map Transportation ................................................................................. 12
       FIGURE 1 ......................................................................................................................................... 14
   3.4 Accomplishing a single national network ......................................................................................... 17
       FIGURE 2 ......................................................................................................................................... 17
4. Conclusions ...................................................................................................................................... 21
   References ......................................................................................................................................... 23
Executive Summary

*The National Map* concept allows for a more effective and regenerative public mapping program but not without some significant challenges to its implementation. The Federal Geographic Data Committee (FGDC) was established by the Office of Management and Budget (OMB) Circular A-16 to establish a coordinated approach to establishing a National Spatial Data Infrastructure (NSDI). However, numerous federal initiatives attempting to consolidate spatial information continue to proceed with fairly poor interagency coordination. Such programs include the Census TIGER enhancement project, *The National Map*, and Geo-Spatial One Stop (GOS). This paper assesses the changing role of federal level mapping organizations and recommends the operational establishment of an overarching authority that can effectively manage *The National Map* and other components of NSDI.

Several organizational issues regarding The National Map are discussed. Management shortcomings dealing with project recognition, leadership, requirements analysis, and implementation planning present the greatest threat to the success of The National Map. Beyond the establishment of an overarching authority, more formal project structure and a phased, cost-benefit approach to *The National Map* development are the suggested remedies.

A long-term road database strategy must be developed in the near term. A sound plan will consider both inter-organizational partnering and technical integration issues. This report recommends that a national, distributed road database be developed. Participants will be required to meet minimal content standards that are firmly founded in a national spatial data requirements analysis. To allow for automated import of local schema data to a global schema, integrating all partner road data, a menu of local system and schema designs will be available for data contributors to select from. The Census TIGER Enhancement project may be developed as the default base network for such a road data base. This strategy avoids redundant road data collection and equitably delegates data maintenance responsibility. Further, data and system interoperability will flourish as local analysis and software development will be shared with little cost.

Several road datasets, including commercial road networks, should be made available via *The National Map* at various subscriber levels. Different access layers to *The National Map* allows for a free, publicly available layer, various levels of subscription for access to commercial or high-cost data sets, and classified access layers for homeland security and emergency management. All federal geospatial data, including all of *The National Map* datasets, should be registered with the Geo-Spatial One Stop (GOS) data portal.

The successful development of *The National Map* will rely on sound management that considers the geographic information systems (GIS) requirements and capabilities across all levels of government. Merging a number of consolidation efforts under one umbrella authority will greatly increase the odds of successful implementation of the NSDI initiative. Integrating public road GIS data and systems is a daunting goal that is a cost prohibitive endeavor to undertake for any single agency. Collectively, however, it is in the best interest of the nation and should be pursued by the administration in the spirit of responsible governance.
1. Introduction

Background

Throughout the 1990s and up to and including the present, the geographic information science industry has rapidly expanded its presence within public governance. While geographic information has long been used in public administration, a number of factors have enabled the increasingly rapid adoption of geospatial technologies within government entities. First, innovations in hardware, software, and communications technology have made geographic information systems (GIS) faster, more interoperable, and capable, with greater data storage capacity. Second, while technological advances have improved GIS operational efficiency, the cost of hardware, such as workstations and networking components, has decreased. Third, advances in spatial data collection techniques have improved the precision, accuracy and, in turn, the value of GIS. Fourth, demand for spatial information and analysis has grown as recognition of its utility in enhancing resource management efficiency is realized. Evidence of this recognition is not difficult to find. Nedvodic-Budic et al. (1999) describe local adoption of GIS technology as a “growth surge”, citing a study that states seventy percent of local governments use GIS. At the federal level, the National Academy of Public Administration (NAPA) (1998) identified twelve federal functions that require spatial data, ranging from national security to economic and community development (USGS, 2001).

Such a rapid surge in the adoption of GIS technology has not come without a price. Many organizations employing GIS have developed and maintained spatial databases mostly independent of other organizations using GIS. This is apparent across all levels of government. One example of this problem is the existence of multiple transportation databases among government agencies - each one designed to fulfill a particular agency mission. The Topologically Integrated Geographic Encoding and Referencing system (TIGER) data support the Bureau of the Census mission; Digital Line Graphs support the basic topographic mapping mission of the United States Geological Survey (USGS); while the Bureau of Transportation Statistics maintains the National Transportation Atlas Database. Further, most every state and many counties have transportation networks that they maintain autonomously. These datasets cover many of the same geographic extents and real-world features, but have been designed and maintained disparately within distinct fiscal budgets. The result is not only redundancy. With budget constraints commonplace, it is very difficult for any one agency to maintain a current spatial dataset for geographies of significant extent. This has been particularly true of the USGS and Census transportation database activities. The USGS states openly that, on average, its topographic maps are 23 years of age (USGS, 2001).

Along with many professionals within the geospatial community, the Office of Management and Budget recognized this operational redundancy in 1990 and released Circular No.A-16 (OMB 2002). This circular, revised in 2002, called for the establishment of the Federal Geographic Data Committee (FGDC) with the purpose of coordinating various spatial data activities among federal agencies. In 1993, the National Performance Review further recognized the importance of spatial information and reported the need to establish a National Spatial Data Infrastructure (NSDI). The vision of NSDI is to create spatial data partnerships across government functions and political levels (Guptill, 1994). The FGDC was placed in charge of developing the NSDI.
USGS and *The National Map*

The evolution of GIS and complimentary technologies coupled with the recognized need for spatial data consolidation has been the impetus for a paradigm shift within the USGS, the DOI entity responsible for maintaining basic, domestic geographic information. Leaders of the USGS geography discipline argue that they must evolve from a map producer to a parent organization that helps to defined national spatial data standards while managing data partnerships with local level governments and private industry where feasible. This vision is encompassed in *The National Map*. It aims to provide the public with basic geographic data across the entire nation through the linking of existing spatial databases at the federal, state, and local levels and privately licensed data where feasible. Within the context of this vision, the USGS (USGS, 2001) will claim responsibility for:

1. guaranteeing national data completeness
2. marketing the availability, and utility of *The National Map*
3. creating and stimulating partnerships
4. integrating, certifying, and quality assurance of data from all participants
5. owning and producing content for *The National Map* where no other suitable and verifiable source exists
6. leading the development and implementation of national geospatial data standards.

The stated content goal of *The National Map* is to provide nationally consistent and integrated topographic map information useful for any arbitrarily defined geographic area. Thematically, content is to include orthophoto imagery, elevation data, hydrography, transportation, structures, government and administrative boundaries, geographic names, and land cover information. The target level of cartographic detail to be achieved is equivalent to or greater than the detail of the USGS standard topographic products.

The arguments for inter-organizational sharing of spatial data are strong. Dueker and Vrana (1995) suggest three classifications of potential improvements from shared GIS initiatives. Although they focus on intra-organizational collaborations, the same potential benefits apply to inter-organizational initiatives such as *The National Map*:

1. Efficiency – pool efforts and allow for data maintenance at a lower per-unit cost,
2. Effectiveness – new and higher quality products, services and analysis and decision making,
3. Enterprise benefits – improved communication and shared knowledge across the expanded community, providing opportunities for further integration and collaboration.

Conversely, the absence of sharing impedes technological advances and the social adoption of GIS (Pinto and Onsrud, 1995). The lack of data sharing forces organizations to expend more resources on data collection and maintenance as opposed to developing analytical applications.
and results. Further, applications for spatial data analysis, once developed, are not easily distributable within GIS communities having extensive system heterogeneity.

Most would agree that there exists a reasonably strong argument for developing a consolidated national mapping program (OMB, 2002; NAPA, 1999; Jensen et al. 1998). There is a great deal of public money to be saved and benefits to be realized in doing so. However, most also realize that there are many obstacles to overcome before The National Map vision comes to fruition (NRC, 2003). What, specifically, are these obstacles facing The National Map and how might they be addressed? This paper aims to describe and recommend strategies for overcoming, what the author sees as, the two most critical challenges facing the implementation of The National Map. Those two major challenges are organizational structure and integration strategy.

Research project focus

The National Map should not be viewed as an extension an existing program or the evolution of an existing governmental entity. Rather, it promises to be a collaborative effort among several governmental and private entities that spans various political and geographic scales. The complexity inherent in such a network mandates the implementation of a thoughtful project design and data integration strategy. The first set of challenges involves organizational constraints that face The National Map. The second set of challenges is technical in nature and addresses the problem of integrating spatial information from different organizations with varying data characteristics and structures.

Many challenges to the successful implementation of The National Map are inherent in institutional bureaucracies. Longstanding institutions have developed and used GIS data for many years to fulfill their own missions and find it difficult to adapt to supporting a broader mission despite acknowledgement of such a need. Daft (1989) demonstrated that bureaucracies are more efficient than open and flexible. Within governmental entities, major changes must typically be initiated with a political mandate before requisite attention and inter-agency cooperation is achieved (Craig, 1995). The National Map is an initiative internal to USGS but its requested participant list reaches far beyond the Department of Interior. So how can the USGS, or the federal government more broadly, learn from past and existing spatial data partnerships to fulfill The National Map promise?

Combining spatial data from various, disparate sources present many potential data integrity problems. Spatial data are not represented consistently among organizations with regard to schema, attribution, or geometric. For transportation networks, there may be significant overlap in geographic extent between neighboring partners with many of the same road segment features represented differently. Geometric inconsistencies across datasets also lead to topological accuracy issues when multiple data themes from multiple sources form a composite map for a particular geography. Unifying the exiting data sources that will comprise a single framework is the process of data integration (Devogele et al. 1998). Developing processes for successful integration in the midst of such heterogeneity is the second major challenge of The National Map.

Both organizational development and data integration issues are pertinent across all spatial data themes for The National Map. Although the implementation strategies suggested herein may be applied project wide, the scope of this study, particularly regarding integration, will be limited The National Map road data.
2. Methodology

To be effective in researching partnership programs, the researcher must play two, somewhat conflicting roles. First, the researcher must remain an objective evaluator. Second, the researcher should also witness the inter-workings of the program by attending meetings and holding informal discussions with the project staff (Ventura, 1995). The author of this report had an opportunity to be a fairly objective observer of The National Map project while also participating in operational level meetings and strategic discussions regarding The National Map implementation. The description of The National Map current strategy and practices are described in this document as the author has witnessed and interpreted from three months as a “semi-autonomous” insider.

The National Map descriptive documentation review

Several different methods were used to fulfill the objectives of this review. The author first needed to more completely understand the argument for developing The National Map. Initially, this included review of the published and unpublished documents dealing with The National Map as well as conversations with USGS employees. Independent evaluations regarding The National Map implementation offered external perspectives on how the USGS may contribute to the NSDI (NRC, 2003; URISA, 2003; USGS, 2002). Beyond specific program evaluations, it was also necessary to attempt to understand the project requirements of The National Map. To complete this mission, several sources were referenced, including the Homeland Security Infrastructure Program publication (NIMA, 2003)¹ and various transportation standard documents.

Interaction with The National Map personnel

Beyond reading articles that relate to The National Map, an effort was made to participate in discussions and meetings with USGS personnel. Opinions and perspective were obtained informally. No transcripts were taken of discussions with personnel. This did not preclude some critical organization issues from being recognized in these discussions. Such issues are addressed later. Frequently, the author was unable to find literature describing The National Map project processes, standards, or justification for key decisions. When this occurred, assistance was solicited from USGS personnel for locating the appropriate documentation. Although, in many cases, no documentation was found, a good-faith effort was made to confirm and acquire or, alternatively, suggest its nonexistence. Overall, discussions with personnel of various functional responsibilities and at various grade levels (GS-7 to GS-15) provided a broad perspective of how The National Map is being implemented within USGS.

¹ This document is currently classified For Official Use Only.
Peer review literature referencing

Various articles and books were referenced in relating The National Map to similar, albeit smaller scale, spatial data collaboratives. An attempt was made to tie the theory learned from spatial data sharing projects to The National Map initiative. Literature of published geographic information sharing, partnering, standards and spatial data integration topics formed the foundation of the implementation strategies presented in this report.

Data evaluation

The content evaluation of potential National Map partnership data was evaluated. Primarily state, local and commercial transportation data sets were compared in the St. Louis and Atlanta metropolitan areas in conjunction with a large USGS research project (Usery et al. 2003). The networks were visually evaluated relative to each other and against recent ortho-photos. The ortho-photos were the assumed ground truth in this analysis due to their high spatial and temporal accuracy and precision. Recommendations for The National Map’s inclusion of any transportation dataset, however, are not based solely on the analysis of these small samples. Rather, the data source recommendations within this report are born from years of experience working with various transportation networks as well as consideration for the organizational and data integration challenges described below.

3. Findings

First, a brief description of current project structure, strategy, and implementation activity is provided. The description is followed by a number of recommended project structure changes that are considered crucial to the long-term viability of The National Map. Next, the challenges presented by large-scale spatial data integration are addressed. A heuristic integration complexity model is introduced as a means for communicating the costs of fully implementing the current National Map transportation strategy. Lastly, a recommended conceptual model for the implementation of The National Map transportation theme is presented.

3.1. Status of The National Map project

The vision of The National Map is made available by the USGS (http://www.nationalmap.usgs.gov/nmreports.html). Therefore, no attempt is made here to further describe the ultimate goal and vision of The National Map. It is important, however, to focus on how the USGS is currently implementing this program. Prior to highlighting some key project shortcomings, a brief description of the operational level implementation of The National Map is neccessary.

The USGS is currently funding The National Map development via the Cooperative Topographic Mapping (CTM) budget. CTM’s mission is to ensure that USGS topographic maps
are available and kept up to date by working with partners in other Federal agencies; in State, county, and local governments; and in the private sector (USGS, 2003). The author does not know exactly how much is being spent on this national mapping initiative, although the Presidential Budget Request for CTM FY 2004 is $74.1 million (USGS, 2003). The Presidential Budget Request for the USGS is available online (http://www.usgs.gov/budget/2004/justification.html). The funding for The National Map is being used to support a variety of current activities, including the development of technical infrastructure, the solicitation of local government participation, and project administration.

Major initiatives for The National Map development have lacked coordination. On July 3, 2003, a memorandum defining the responsibilities of The National Map Business Model Team was distributed to the Geography Senior Staff within USGS. This team is tasked with maturing the draft implementation plan and coordinating efforts between the various systems development design teams. Various technical design teams include systems architecture, Web Mapping Service (WMS), and feature data model development. The formation of the Business Model Team had not yet positively impacted the operational level of The National Map development as of early August, 2003. While detailing each of the above team activities is beyond the scope of this project, a description of the WMS strategy, essentially the spatial data management system of The National Map, is provided in section 3.3. Essentially, a cross-reference table relates local data schemas to The National Map schema. This enables data access to a distributed network of data within a single map interface.

The solicitation for local government participation in this data-sharing concept is well underway. USGS Mapping Partnership Offices (MPOs) are being opened across the country where they will be urging state and local government agencies to share their data with The National Map. Many state and local governments are providing data to the USGS. However, the spatial extent, theme and quality of the data vary greatly. It is too early to tell how pervasive and persistent cooperation with local governments will be.

Along the administrative front, The National Map management is somewhat of an oxymoron. All of the strategies described above, and presumably many more, are taking place while project administration for The National Map is still scripting its plan for implementation. Pilot projects are being conducted and thematic data standards teams have been formed. Each team exists in relative isolation, with no formal communication between them. Consequently, little has been produced from these teams thus far, and many concurrent activities seem to be vaguely coordinated. Below, the organizational obstacles regarding the implementation of The National Map and, more broadly, the NSDI are detailed further below with recommendations for overcoming them.

### 3.2. Recommended project changes

While it is commonly recognized that both technical and organizational impediments to spatial data sharing and integration exist, much research suggests that the organizational obstacles are more difficult to overcome (Croswell, 1991; Masser et al 1995; Nedvodic-Budic et al. 1998). Masser and Campbell (1995) highlight some key obstacles to GIS sharing:

1. variation in participant priorities,
2. variation in GIS experience and technical ability,
3. differences in spatial data handling skills,
4. disagreement among participants regarding data openness, leadership, data standards, equipment and training.

Given the complexity of a national spatial data integration program, these issues are very difficult to surmount. Meredith (1995) shows that the greater the number of participants in a data sharing program, the greater the organizational complexity. In addition, an inverse relationship between the interdependency of sharing organizations and the likelihood of project success has also been noted (Azad et al. 1995).

Perhaps, recognition by the USGS of these organizational impediments to data sharing constitutes their argument for the need of a zero-mandate policy with state and local governments. However, such an under-bound network of partners may produce a National Map product of such low quality that its cost may not be justified. Integration issues related to this bottom up approach are discussed at length later. Nonetheless, the above organizational impediments, when considered with The National Map and NSDI visions, present some fundamental project realizations:

1) The cost of spatial data sharing is significant.
2) Successful development of The National Map is dependent upon an up front understanding of The National Map stakeholder community needs and capabilities.
3) Long-term implementation must find a balance between mandated compliance with standards for participants and local control of data management.

Project recognition and consolidation

Recognizing the importance of The National Map and its relation to other federal geospatial initiatives is critical for the efficient management of spatial data. The participant and potential stakeholder community of The National Map reach far beyond the bounds of USGS authority (NRC, 2003). In fact, there is considerable overlap with other federal initiatives to consolidate spatial data management. GeoSpatial One-Stop (GOS) and the Census TIGER Enhancement program are two other major geographic data collaborative efforts underway. While the USGS implementation plan (2003) recognizes a need to coordinate efforts between GOS and The National Map, the Census TIGER database development effort remains fairly distinct from the USGS activity. For example, while Harris Corporation is trying to develop means for developing partnerships and integrating local transportation network geometry with Census geometry, the USGS is independently soliciting local data sharing and researching integration strategies for The National Map. Further, the Census and USGS have independently maintained a database of state and local government organizations that share data with the Census and the USGS respectively. This has resulted in both the USGS and TIGER soliciting the same information from a single, local government, illustrating unnecessary data sourcing duplication.

Serious consideration should be given to consolidating The National Map activities with other geospatial initiatives, such as the GOS data portal development program and the TIGER Enhancement program. It is important for the administration to recognize that all federal level geospatial data collaboratives should reside under the umbrella of an overarching NSDI implementation authority. As Ventura (1995) argues, an overarching body must be recognized and granted clear authority over the participating organizations.
The FGDC would be the government entity best suited for meeting the managerial
demands of such operations. However, the implementation committee should be comprised of
appointed officials, as the committee members must serve without any perceived bias toward any
one agency (Nedvodic-Budic, 2001). This overarching body must also hold executive authority
and accountability over all NSDI implementation projects, including project funding
mechanisms as well as management should (Ventura, 1995). Presently, some FGDC members
are not actively participating in its activities. Likewise, some federal agencies are not employing
FGDC’s standards. A National Academy of Public Administration panel found that FGDC’s
strategy for implementing NSDI is not reflected in agency strategic plans and annual
performance plans developed under the Results Act (NAPA, 1999).

**Project leadership**

Following the identification of an overarching authority, *The National Map* leadership
must first establish a project management structure that defines and delegates priorities. This is
critical. Poor coordination and project leadership is the leading factor in the termination of data
sharing programs (Nedvodic-Budic et al. 2000). Currently, there is a lack of formal project
structure within USGS for making efficient progress toward the ambitious National Map vision.
And formality will become increasingly important as USGS Geography expands the partnership
program with other government entities (Johnson et al. 2001). This lack of formality has allowed
for confusion in the responsibility of project personnel and the urgency of assigned tasks. Such
shortcomings were witnessed regularly as theme standards teams fumbled past vague deadlines
without notice by more senior project staff². Notwithstanding a few committee and inter-agency
meeting notes, there is an absence of published procedures for developing *The National Map*. If
such documentation does exist, it resides in too few hands. In three months as a USGS contract
employee, the author, with the assistance of GS15 USGS personnel, was unable to locate any
procedural or project requirements documentation for *The National Map* planning and
development (Canfield, 2003 pers. com.; Domaratz, 2003 pers. com.).

**Project Requirements**

Once clear project authority is assigned and leadership established, making the transition
from the grand vision statement to a feasible implementation plan will require a thorough
evaluation of the geospatial community needs, at all levels of government. Understanding what
data, structures and procedures are required by all classes of participants prior to implementing
hands-on development plans provides the best opportunity to develop the correct implementation
strategy. As an example, the Australian State of Victoria successfully implemented a partnership
program whereby the State provides the local governments with basic map data and the local
governments provide the state with parcel data. After years of floundering, the first step toward
success culminated in one of the largest requirements studies ever conducted by an Australian
State (Jacoby et al. 2002). Currently, data standards teams for *The National Map* have been
formed and tasked with detailing the content standards for each of the eight geospatial themes.

² The phrase “senior project staff” is used in lieu of “project management” due to the lack of recognized or explicitly
designated project management personnel.
However, they have no documented data requirements to work from nor do they have a mechanism for coordinating their efforts. Section 2.1.5 of the current implementation plan draft reads:

“2.1.5 Identify Customer Requirements
Section is under development.”

When one considers that The National Map is expected to be a voluntary network of participants, the project requirements phase gains further importance. In the absence of a legal or political mandate, long-term participation in an inter-organizational data-sharing program requires the perception of mutually reinforcing benefits (Azad et al. 1995; Cook, 1977). This argument is known as exchange theory, and it is likely the most popular reasoning for organizational cooperation within the public sector (Azad et al. 1995).

With the notion of exchange theory in mind, it follows that a comprehensive requirements study that includes representatives of each stakeholder agency at the federal level, and primary spatial data functions at each political scale is paramount. This goes beyond inter-agency meetings and coordination. Rather, a mandated procedure for decision-making that requires input across all significant domestic federal mapping and geospatial analysis functions is recommended. Private sector participation should be encouraged as well, making use of commercial GIS expertise in spatial data collection, management, and distribution.

**Project implementation**

A phased implementation plan should be developed using an appropriately detailed requirements document. The implementation plan should unambiguously state the mission of The National Map and specifically define its bounds. Recent criticism of USGS published documents on The National Map, such as the National Research Council’s “Weaving a National Map” (2003), have frequently questioned the practicality and lack of strategic clarity for achieving The National Map vision. This is due, in part, to the ambiguous language that is used to describe The National Map. For example, what does seamless, integrated or consistent spatial data mean with regard to The National Map. But more importantly, the USGS has not publicly addressed how it will tackle the major organizational and technical impediments to such a program. For example, how will information be made seamless? How will The National Map be standards based while not mandating any level of data standardization among state and local participants? These are two of the basic implementation questions that should be thoughtfully considered prior to commencing major implementation programs. In the absence of clear language and a sufficiently detailed plan for attaining the goals of The National Map, confusion and inefficient project development within and outside of the USGS will persist.

It must be acknowledged that many of these project issues have much to do with the lack of political mandate for The National Map project. Recognition and independent funding for The National Map development, coming from the overarching authority, described above, will provide the necessary impetus for inter-agency collaboration and project organization.

Beyond the need for reorganizing bureaucracies in support of federal GIS consolidation, there is a need to make spatial data more interoperable. Integration of longstanding, disparate data sets is the major technical hurdle of any such effort. There are several levels of systems
integration that apply to inter-organizational spatial data sharing and consolidation programs. The following list presents the forms of systems integration offered by Deuker and Vrana (1995, 153). They are listed in increasing order of comprehensiveness.

Taxonomy of systems integration:

1. **Data Integration.** The process of unifying existing data sources representing some or all of the same entities into a single framework of data (Devogele et al 1998; ESRI 2001).
2. **Applications Integration.** Bundling separate applications into larger, more comprehensive applications.
3. **Functionality Integration.** Consolidating applications that serve different organizational functions with the purpose of enabling the support of a broader range of functions.
4. **Organizational Integration.** The integration of various units within an organization to achieve a common objective. This could mean the combining of functions, applications, personnel teams, and office space.
5. **Mission Integration.** The integration of organizations for the purpose of achieving cooperation towards a shared vision. This is principally a political and cultural level of integration.

Any legitimate long-term plan for *The National Map* integration must consider the subsequent levels of the integration taxonomy. The public geospatial community should strive for the maximum level of data interoperability while also fulfilling their distinct obligations. The remainder of this project report will focus on the challenges of integration and a recommended, phased approach for the development *The National Map* transportation theme.

### 3.3. Integrating *The National Map* Transportation

There are two distinct types of integration: horizontal integration and vertical integration. *Horizontal integration* is defined as the joining of spatially adjacent datasets with the purpose of generating expanded geographic coverage of for a particular data theme. This term is intentionally broad, as project requirements determine the specific criteria for horizontal integration. For example, a data integration project may have geometric edge-matching of border road segments as the sole criteria for two datasets to be horizontally integrated. A more rigorous application may require that all participant data be interoperable, meaning road segments have unique object identifiers, road paths be consistently attributed, and have geometric edge-matching in order to be considered horizontally integrated.

*Vertical integration* refers to the cartographic overlay of data themes that originate from more than one source. Topological integrity across themes and data models is the goal of vertical integration procedures. As with horizontal integration, vertical integration success is measured by the error tolerance threshold defined in project standards and application requirements. Although there is no simple solution to either type of integration, horizontal integration may present the greater challenge to the transportation theme, due to the variety of datasets to be integrated. While this section will allude to vertical integration issues, its focus is on the horizontal integration of transportation data.
There exist three main strategies for accomplishing spatial data interoperability offered by Devogelle et al. (1998). The first approach is to integrate the data manually, specifying the data from participating databases that is to be merged with the global application. Global application processes receive the component data and synthesize the information to meet the global application requirements. A second approach to interoperability is through the application of standards. Standardization can be applied to both data models and schemas. Essentially, standards facilitate data exchange among systems. With a wealth of valuable information already existing, however, the problem of converting present, unstandardized, data to a standard format remains a problem with this approach. The third strategy for achieving interoperability is to develop a software system that ties together existing data model and schema designs. This is referred to as a federated database system (FDB) approach. It requires that all local schema differences be resolved via a global, virtual schema (Devogele et al. 1998, 336). This would be a daunting task for The National Map considering the number of participants. A federated system may be developed with either a bottom-up (local translated to global schema) or a top-down (global schema mandated to local) approach (Laurini, 1998).

What strategy is currently proposed for achieving data interoperability for The National Map? As briefly described in section 3.1, The National Map currently relies on a bottom-up strategy for data integration that is most closely related to a FDB approach. The current strategy for dealing with horizontal integration of transportation is to develop automated and semi-automated techniques for linear feature edge-matching. From a data attribution perspective, USGS staff will translate local schema designs to The National Map global schema via maintenance of a cross-reference table. The benefits of this strategy include limited disturbance of local participants’ existing processes. Any standards that are applied with this approach are applied at the overarching level. This translates into less of a need for USGS to incentive local participation, monetary or otherwise.

There are many disadvantages to the current USGS integration approach, however. A bottom-up approach is much more complex than a top-down approach due to the heterogeneity of existing spatial databases (Laurini, 1998). Responsibility for tying together partner data is left to the overarching authority, along with the cost for doing so. Of primary concern is that this process does not address the problems of inconsistency among the various jurisdictions. An enormous amount of work is left up to the organizing entity, i.e. the USGS with some major issues left unresolved. Figure 1 illustrates the basic philosophy of this mapping service structure and some of its limitations.
The above figure illustrates some classic complicating factors with the bottom-up federated approach. First, a partnership is established with a local authority. Second, each stream of information from the partner’s WMS needs to be interpreted and cross-referenced to The National Map schema before it can be displayed. A complete integration of schemas, where local schema discrepancies are resolved, will be forgone in The National Map due to the volume of participants. The cross-reference table above shows how the local semantics will rarely match the global semantics. There will rarely be a 1:1 match of feature classes between the local and global data structures. Further, the data layer containing major roads for any one local participant is most likely unique from all other participants when considering the various systems, cartographic representations, and definitions for a major road. With the number of partners envisioned coupled with the heterogeneity of mapping schemas and systems, maintaining thousands of partnerships will be awfully burdensome.

Beyond the classic schematic integration issues, discrepancies specific to geographic information presents further challenge to a distributed database design. The typology of spatial database issues below is based on a list of issues presented by Laurini (1998, 380).

1. Diversity in spatial representations,
2. Diversity in global projections,
3. Diversity in values for the same items located at different sites
4. Diversity of spatio-temporal sampling,
5. Variability of definitions over time and space,
6. Discrepancies in co-ordinate values,
7. Discrepancies in boundary alignment (zonal fragmentation),
8. Variability in content quality,
9. Variability in data maintenance procedures,
10. Discrepancies in spatial metadata
Given the complexity of integration issues, it is difficult to envision how the current strategy can be employed over the entire nation and meet the transportation theme requirements of *The National Map* and NSDI stakeholders. A phased approach to project implementation is recommended. As mentioned earlier, a formal process, including roles and documentation, must permeate each phase.

**Phase 1: Determine transportation requirements and national capabilities**

A sufficiently detailed requirements and capabilities study should be the first initiative in developing the transportation theme. The product of such analysis would inform decision making in the subsequent phases. It is important to involve a wide range of federal agencies. As of 1999, as many as 40 US agencies were using geospatial data (NAPA, 1999). State and local government must also be represented in this process. Questions regarding future information needs as well as the technical, financial, and human capital assets that exist at all levels of government must be examined. The geographic variability of such needs and capabilities should be considered in formulating sound alternatives for the transportation theme. Some of the fundamental transportation network questions that need answered across all government levels in this phase are listed below.

**Requirements:**
1. What is the thematic priority of spatial data set for governance at each political level?\(^3\)
2. Generally, how is transportation data used currently (e.g. routing, geocoding)? In the future?
3. Who are the users?
4. What information needs to be related to the transportation network (e.g. parcels or structures)?

**Capabilities:**
1. What is the current data quality?
2. What processes, procedures, and personnel maintain it?
3. How is transportation data development and use financed?
4. What is the system design and schema of all federal transportation data and a sampling of state and local data?
5. What transportation data partnerships exist among the community?
6. Who is willing to share data and/or pool resources?

**Phase 2: Develop data integration alternatives**

Once public transportation data needs and management has been identified, data integration alternatives should be developed. In this phase, distinct, practical strategies for meeting the transportation requirements should be conceived and put forth. However, these strategies must also be informed by the constraints illuminated in the capabilities study.

---

\(^3\) The purpose in asking this question is to determine how the concept of base geospatial data varies across political levels or agency missions. Many cross-level data collaborative efforts have shown cadastral information to be the most critical data for local governments (Jacoby et al. 2002; Nedvodic-Budic et al. 2000). Hence, the priority of topological relations across themes may vary between political levels.
Phase 3: Cost-benefit analysis of alternatives

A cost-benefit analysis must ensue the development of alternative approaches. To assist in the evaluation of cost for any particular transportation strategy, a heuristic model that quantifies the complexity of a given integration strategy for The National Map transportation theme should be developed. The following variables provide a basis for such a model.

Transportation integration complexity theory:
1. The length of coincident participant boundaries,
2. The number of participant schemas to be unified to The National Map global schema,
3. Average linear road network feature density (e.g. 2 roads per linear transect mile),
4. A rating of participant interdependency

A few assumptions must be met prior to applying the model:
1. All data partners must meet spatial precision and accuracy requirements with their networks.
2. All participant data is available – i.e. there is no variability in data acquisition cost.
3. Partner boundaries are only considered distinct from one another if they do not coordinate edge-matching and efforts. If two county level data partners, for example, reference each others’ network for edge-matching, then the border between the two counties would not contribute to the length of coincident participant boundaries (variable 1).

In addition to quantifying the difficulty of data integration, the cost of maintenance for each alternative must be estimated. A solution that does not provide a means for participants to reference one another’s data holdings will require repeated, periodic edge-matching. Local schematic changes are another cost that should be considered with each alternative strategy. Over time, a strict, centralized standard would cost much less to maintain than an unstandardized, decentralized data network. At the same time, the initial cost of a centralized standard would be high as local participants modify their systems to achieve compliance.

The benefits for each alternative should be reviewed against the costs. What will the resulting system or suite of products have to offer government and the public? The benefits will be difficult to quantify. One way to evaluate each approach is to list the types of applications that would they would make possible. For example, which alternatives will provide a coast-to-coast geocoding support capabilities? How much data maintenance money will be saved across all participants? Although much of this work is no more than an educated guessing game, at least it is educated to the greatest extent feasible. And relative advantages among the various alternatives, weighed against their costs, may help to illuminate the best approach.

Phase 4: Physical implementation

The overarching body should make the final decision on which alternative to implement. Once chosen, a detailed implementation plan should be developed and subsequently rolled-out. It is vital that the implementation be actively managed and modified when necessary (Budic 2000). There is no empirical understanding of geographic information partnership programs of this
magnitude, across political levels and organizations types. Therefore, it is imperative to efficiently coordinate systems development, partnerships and research for *The National Map* in order to ensure that the cost of implementation may be recouped by accrued benefits within a reasonable timeframe.

### 3.4 Accomplishing a single national network

The long-term solution to integrating national street centerline data is to develop a distributed database of national extent that meets the scale requirements of local governments. Jensen et al. (1998) calls for the need to establish a centerline database accessible to all levels of government. Such a database could be maintained at the local level with no duplication of effort at the federal level. Centralized control of standards, informed by all stakeholders, would constitute the mechanism for information exchange and interoperability.

Figure 2 illustrates the suggested, long-term conceptual data flow for *The National Map* Transportation. Although such a program cannot be implemented immediately, the overarching authority of NSDI needs to develop a plan for achieving such a distributed non-redundant data maintenance program. The phased project approach described above should be used to determine the means for achieving this vision.

**FIGURE 2**

![Conceptual Data Flow](image)
The schematic is detailed by stage number labeled on the data flow diagram. Generally, the flow of information is from left to right. Participating data contributors maintain data on the left, committing it to a centralized road data repository via a translation process, and a cartographic template application for *The National Map* rendering on the right of the diagram.

**Stage 1: Data Partners**

The first stage of the data flow recognizes the participating data providers to *The National Map* transportation theme. Under this design, each participating agency will have a menu of local system-schema designs to migrate their data too. The menu will be a list of systems and schemas derived from the phase 1 requirements study, identifying the most common data structures among the data providers. Data will reside locally with each participating organization. Eventually, the majority of transportation data content within *The National Map* primary network will come from local organizations. And minimum content standards and capture conditions must be adhered to by all data providers.

Census TIGER files, it is recommended, should provide the base network for transportation. Many local government transportation networks are based on TIGER files. Therefore, many local GIS agencies are familiar with the format and history of TIGER data. In addition, the TIGER Enhancement project is actively soliciting local participation in improving the spatial accuracy of its network. Although no network maintenance plan beyond the 2010 Census has been publicly discussed, it is recommended that a decentralized, ongoing data development effort be initiated.

The National Forest Service (NFS) and The National Park Service (NPS) represent other federal agencies that have unique road network requirements. The Forest and Park Services need forest fire roads mapped. Since this demand is unique to them, they will then maintain public land fire road geometry and attribution. **The objective is to have each agency fund the data development effort proportional to their respective use of it.** Data flows from each distributed database through a translation process and to a national Road DB. Local reference to neighboring partner data for edge-matching may be accomplished through a view of the Road DB.

**Stage 2: Schema Translation**

With each participant storing transportation data within a known (via metadata) system and schema, automated data translation procedures may be developed for importing all participant data to a centralized data base. Data from each provider will be committed periodically, based on demand for each partner’s data, the frequency of network change, and schematic translation processing capacity constraints.

**Stage 3: Central Transportation Database**

The Road DB will hold all geometry, primary key assignments for road segments and nodes, as well as minimal attribution for cartographic display. However, the majority of the network attribution will remain distributed. Each partner will be periodically provided with available primary key sequencing files, allowing for complete functional autonomy from the Road DB during database maintenance. Therefore, all data committed into the Road DB will already have had primary key assignments for nodes and road segments. The Census Master Address File will be related to the Road DB similar to the manner in which it is currently related.
to the TIGER files. Network continuity validations, nation-wide quality assurance, or other analytical procedures may be developed and applied to the Road DB by USGS researchers or other agencies interested in the continuity of the national Road DB.

Stage 4: Cartographic Rendering

A default cartographic template will be maintained by the USGS or government contractors for the purpose of displaying the Road DB via *The National Map*. Custom templates may be developed for various purposes by any interested party. The template will also incorporate other data sets to be displayed on *The National Map*, including the other seven NSDI themes and restricted (non-public) access data. Maintaining the cartographic template, and WMS more broadly, may be a function the federal government should consider outsourcing to the private sector. There is a wealth of experience and success in the private industry that should be tapped.

Stage 5: The National Map Product

*The National Map* will provide the cartographic window to The Nation’s transportation data providers. All national map data will be viewed through this WMS. Again, a commercial applications service provider may be contracted to host *The National Map* site.

Thus far, only the primary, public transportation data has been recognized within this data flow. However, *The National Map* will have multiple access levels, from unrestricted (public) to classified levels for national security and emergency management access only. There will undoubtedly be a need for other, more robust, transportation networks to support sophisticated routing applications. Although recent discussion within an inter-agency transportation theme team recommended that TIGER have routing attributes added to it by any agency that is interested in routing on *The National Map* data (Canfield, 2003 pers.com.), the reality is that such an effort would be cost prohibitive. A network that supports routing applications not only requires attributes such as direction but also requires topological design considerations that are beyond the framework of the TIGER data model.

The public sector would be wise to leave routing database development to commercial data providers. Organizations may be granted various levels of access to these networks via a *National Map* subscriber service. Of course, the commercial data venders such as NavTech, GDT, or TeleAtlas would receive royalties for the use of their data. Government should be able to take advantage of private routing and geocoding software that has been developed around the commercial networks. Applications relating event locations on the Road DB to a commercial network, such as NavTech, can be developed by utilizing the comparable spatial precision and addressing attribution of the two road networks.

Stage 6: Other Data Theme Relations

Other data themes or data sets may be explicitly or implicitly related to the Road DB. The dotted line in figure 2 signifies that other data sets may have foreign keys to the Road DB or visa versa. Records within the Structures model, for instance, may require a foreign key to a road segment identifier. Explicit linkages across themes must be considered in the design of *The National Map* as a whole. This stresses further the importance of a coordinating standards development effort among data themes. As discussed above, other transportation data may be indirectly related to the Road DB via network conflation applications. Spatial data sets should be
available for view within a *National Map* access layer (public or restricted) and should also be registered with the Geo-Spatial One Stop portal (GOS).

**Stage 7: Geospatial One-Stop Relation**

It is recommended that *The National Map* be coupled with GOS. *The National Map* would serve as the data view of information within GOS. Users should have the option of identifying participant data sets for download via *The National Map* map interface or by directly selecting the data sets listed with GOS. The point being is that there is a 1:1 match between *The National Map* directory of data sources and the GOS directory of data sources.

The alternative described above should be considered by *The National Map* transportation team. It is unlikely, however, that this approach would be adopted without political elevation of *The National Map*. That said, many shortcomings of the present approach the USGS is taking to develop *The National Map* road data are addressed with the conceptual approach described above. Some of the benefits are listed below:

1. Primary key attribution of road segments is resolved with the above approach. Unique identification of road segments and nodes within a data structure will be a requirement of American National Standards Institute’s (ANSI) transportation standard (FGDC, 2003).

2. Maintenance of a global schema is greatly simplified when a moderate level of schematic standardization is applied to the participant network. There is no longer a need to constantly monitor each participant’s data for changes in schema design or capture conditions. With the present USGS approach, the technical data structure may stay the same, however the local data capture conditions may change. Without a communication mechanism to *The National Map*, such a change would go unnoticed. For example, a local government may decide to no longer map alleys, driveways, and unpaved roads due to local budgetary constraints. The database structure may remain the same but the maintenance procedures that dramatically affect content would change. Such procedural changes may be invisible to the USGS without a coordinating effort from the local data provider. This report recommends that minimal schematic design and capture conditions be adhered to by all participants, allowing for both a temporally and spatially consistent Road DB.

3. Inconsistent, unstandardized content may require customized and rapidly evolving integration procedures for various partners. Unless data is edge-matched in the road data bases, as recommended, integration procedures must be applied with each data update.

4. Partnership interoperability development will not be inhibited by the Road DB design. Whereas the heterogeneity of current systems, capture conditions, content quality may make it difficult to perform spatial or network analysis across partnership datasets. Existing software for TIGER and commercial datasets will be utilized with this suggested strategy, for instance. This effectively decreases the potential benefits of a national road data layer. With some level of standardization, however, applications developed by one municipality could be more easily marketed to other municipalities – considerably decreasing the total public cost of system development.

5. The proposed strategy establishes neither an overbound or underbound network organization (Golembieski, 2003 pers. com.). Without any standards being established at the participant level, massive confusion and inconsistent quality will characterize *The National Map* implementation. At the same time, it is recognized that too much top down
control will increase the participatory cost of local governments, perpetuating redundant, public GIS development. The standards development process must recognize the diversity of local data management systems and refrain from mandating how to maintain spatial data. Rather, standards should be data focused, merely addressing basic schema design and minimum database content.

A great deal of consideration was given to the use of commercial data as a primary road data source. After all, data providers such as NavTech maintain spatially precise databases with rich attribution and nation-wide coverage. There would be no need for horizontal integration if a single data provider is selected to provide road data for the entire nation. But the competitive market is the motivating factor that drives commercial data providers to create such robust data sets. With the adoption of a single commercial provider for roads, and its subsequent integration with other The National Map themes and applications, the competitive incentive for the commercial data provider would diminish over time. Consideration for the use of commercial road data as a primary provider should not be ruled out in the short run, however, and the competitive market for GIS road data must be studied in greater detail before confirming the validity of this perspective.

4. Conclusions

Paramount to The National Map implementation is an administration level recognition of the need for an overarching authority for consolidating the Nation’s geospatial data. The vision of a national base map offered by the USGS merits the full attention of the geospatial community. However, the successful implementation of this national mapping program is beyond the control of the USGS. Procuring data, developing appropriate data maintenance strategies, and coordinating geographic information sharing partnerships encompasses more than twelve federal agencies (USGS, 2001), forty federal functions (NAPA, 1999), and countless state, local and private organizations. Further, The National Map vision is very closely related to other federal geospatial consolidation programs, such as Geo-Spatial One Stop. A viable long-term approach to The National Map development will require the direction of an objective, accountable, overarching body that will establish geospatial product requirements and develop a national geospatial consolidation plan.

Current project management structure in the research and development of The National Map is lacking within the USGS. Premature activity regarding system design and development are abound as The National Map vision has begun large implementation efforts prior to developing a sound implementation plan. A requirements assessment for The National Map stakeholders is far from complete and with no apparent beginning. Overall, there is a lack of adequate formal project structure within USGS for the roll-out of such a complex initiative. This is partially due to institutional inertia among the stakeholder community, a lack of congressional mandate for The National Map, and no clear interagency authority.

The long term issues of integration and interoperability must be addressed at the outset of The National Map development. With data requirements as a guide for implementation, an overarching authority must consider the costs and benefits of various alternatives to road data strategies. A heuristic integration complexity model is recommended to assist in the cost
assessment of each strategy. The total estimated cost of each strategy should be weighed against the benefits of their respective results.

It is further suggested that a viable, long-term approach to The National Map road data includes the development of a distributed road database. Such a database would be maintained by a combination of local and national data contributors. Each contributor should be responsible for maintaining road data within its geographic jurisdiction while conforming to minimal standards that provide for national base data consistency. A phased approach to implementation is suggested, with resources focused solely on the development of a sustainable, long-term system.

The National Map can be implemented successfully with more thoughtful consideration of the broad geospatial community. Careful attention must be paid to public spatial data demands across governing levels, not merely among federal agencies. Its place within the NSDI must be clearly defined and managed at the appropriate executive level. If properly implemented, The National Map will serve The Nation well through more effective governance, higher data quality, and an unprecedented window to the world for the public.
References


Canfield, D. 2003. Personal communication. Deputy Branch Chief, USGS Mid-Continent Mapping Center, Rolla, MO.


Golembieski, R. T. 2003. Personal communication. Professor of Public Administration, University of Georgia, Athens, GA.


