The Evolution of SDI Geospatial Data Clearinghouses

Maurie Caitlin Kelly
The Pennsylvania State University, USA

Bernd J. Haupt
The Pennsylvania State University, USA

Ryan E. Baxter
The Pennsylvania State University, USA

INTRODUCTION

Geospatial data and the technologies that drive them have altered the landscape of our understanding of the world around us. The data, software and services related to geospatial information have given us the opportunity to visualize existing phenomena, to understand connections, and to address problems from environmental management to emergency response. From the ever-present Google Earth images we are shown in our televised weather reports to the 3D flyovers of war zones on the news, geospatial information is everywhere. In the decade or so since U.S. President William Clinton set the stage by announcing the establishment of the National Spatial Data Infrastructure (NSDI), the concept of the geospatial data clearinghouse has shifted dramatically to fulfill the increasing need to streamline government processes, increase collaboration, and to meet the demands of data users and data developers (Clinton, 1994). The announcement of the NSDI gave birth to a Global Spatial Data Infrastructure (GSDI) movement that would be supported by a network of SDIs or geospatial data clearinghouses from local, state, and national levels.

From this point on, the evolution of the geospatial data clearinghouse has been rapid and punctuated with challenges to both the developer and the user. From the earliest incarnations of these now pervasive resources as simple FTP data transfer sites to the latest developments in Internet Map Services and real time data services, geospatial data clearinghouses have provided the backbone for the exponential growth of Geographic Information Systems (GIS). In this section, the authors will examine the background of the geospatial data clearinghouse movement, address the basic phases of clearinghouse development, and review the trends that have taken the world's clearinghouses from FTP to Internet Map Services and beyond.

THE SPATIAL DATA INFRASTRUCTURE MOVEMENT

No discussion of SDIs and geospatial data clearinghouses would be complete without a brief introduction to the history of the movement.

The growth of geospatial data clearinghouse movement can trace its origins to the spatial data infrastructure initiatives of the 1990s when spatial data sharing began in earnest. In the United States an effort to organize spatial data and develop standards for sharing data began as the NSDI. First envisioned in 1993, the concept of the coordinated data model set forth the ideas and goals of widespread sharing of data and resources (National Research Council, 1993). By 1995, the United States had developed a plan for data sharing and established a gateway by which participants could register their metadata holdings through a centralized source (FGDC95). Sharing data through this gateway required developing metadata to an accepted standard and utilized the Z39.50 protocol—both of which will be described in the next section.

The spatial data infrastructure concept as it has evolved has, at its core, the premise that sharing data eliminates redundancy, enhances opportunities for cooperative efforts, and facilitates collaboration. In addition, the SDI movement also has two additional advantages. First, it allows a more effective and efficient interaction with geospatial data and, second, it helps to stimulate the market for the geospatial industry

(Bernard, 2002). The general approach to developing an SDI is to first understand how and where geospatial data is created. Most SDIs or geospatial clearinghouses base their first level data collection efforts on framework data (FGDC95). Framework data is created by government agencies—local, state, federal, or regional for the purpose of conducting their business such as development and maintenance of roads, levying taxes, monitoring streams, or creating land use ordinances. These business practices translate themselves, in the geopspatial data world, into transportation network data, parcel or cadastral data, water quality data, aerial photographs, or interpreted satellite imagery. Other organizations can then build upon this framework data to create watershed assessments, economic development plans, or biodiversity and habitat maps. This pyramid of data sharing—from local to national—has been the cornerstone of the original concept of the SDI and considered a fundamental key to building an SDI (Rajabifard & Williamson, 2001).

The SDI movement now encompasses countries and regions all over the world and is now considered a global movement and potential global resource. Many countries maintain now clearinghouses participating in regional efforts. One effort along these lines is the GSDI (Nebert, 2004). The GSDI, which resulted from meetings held in 1995, is a non-profit organization working to further the goals of data sharing and to bring attention to the value of the SDI movement with a particular emphasis on developing nations (Stevens et al., 2004). Other projects including the Geographic Information Network in Europe (GINIE) project are working toward collaboration and cooperation in sharing geospatial data (Craglia, 2003). As of 2006, there were approximately 500 geospatial data clearinghouses throughout the world. The activities of the clearinghouses range from coordinating data acquisition and developing data standards to developing applications and services for public use with an average operating cost of approximately € 1,500,000 per year (approximately $ 1,875,000) (Crompvoets et al., 2006).

EVOLUTION OF SERVICES AND ACCESS IN THE GEOSPATIAL DATA CLEARINGHOUSE

There are several developmental phases that geospatial data clearinghouses engage in to become fully operational and integrated into a larger SDI, e.g., data acquisition and documentation, data access and retrieval capabilities, storage architecture development, and application development. These phases can be sequential or can be performed simultaneously but all must be addressed. It is important to note that technology, both internal and external to the clearinghouse, changes rapidly and therefore any clearinghouse must be developed to be dynamic to meet the changing nature of the technology and the changing needs of its users. Each geospatial data clearinghouse also must address the particulars of their organization such as available software, hardware, database environment, technical capabilities of staff, and the requirements of their primary clients or users. In some cases, clearinghouses have undertaken an effort to develop user requirements and assess needs prior to implementation of new services or architectures. The user needs and requirements assessment addresses all phases of the clearinghouse from both internal and external perspectives and provides the framework with which to build services and organizational capability (Kelly & Stauffer, 2000). Within the requirements phase, examination of resources available to the clearinghouse must be determined and if inadequate, acquired. There is little doubt that the key to success relies heavily on the resources of the geospatial data clearinghouse and its ability to store and provide access to large datasets and thousands of data files (Kelly & Stauffer, 2000). Another equally significant component of building an SDI is identifying how the resource will support activities in the region. The clearinghouse can bring together disparate data sets, store data for those organizations that are unable to store or provide access to their own information, and can offer access to data that crosses boundaries or regions to enable efforts that are outside traditional jurisdictions of agencies or organizations (Rajabifard & Williamson, 2001).

Metadata

The key component to any geospatial data clearinghouse is geospatial metadata. The metadata forms the core of all other operations and should be addressed in the initial phase of clearinghouse development. The Federal Geographic Data Committee (FGDC) developed its initial standards for geospatial metadata in the mid 1990's. This standard, which is used as the basis for metadata in geospatial data clearinghouses today is re-