Spatial Data Infrastructures

Clodoveu Augusto Davis, Jr.
Pontifical Catholic University of Minas Gerais, Brazil

INTRODUCTION

Spatial Data Infrastructures (SDI), also known as Spatial Information Infrastructures (SII), are a set of policies, technologies and standards that interconnect a community of spatial information users and related support activities for production and management of geographic information (Phillips, Williamson, & Ezigbalike, 1999). SDI reduces redundant effort and lowers production costs for new and existent datasets through interoperable information sharing, providing neutral means to access geographic data. Multiple information providers, commercial or public, may cover various interests and compete among themselves for clients.

SDIs present several challenges, at various levels of interaction. First, there is a societal and organizational level. Partners in a community should have convergent interests, agree on common rules, and be able to use information produced by others. Such agreements are not easy to achieve, and usually require long-term commitments. Within public organizations, it is usual to think in transnational terms, between national mapping agencies, but intranational relationships are also important.

Second, there are standardization issues. Guiding the technology standardization and defining the key elements for SDI, the Open Geospatial Consortium (OGC) has proposed a number of standards, through a framework called *OGC Reference Model* (Percivall, 2003).

Third, there are concerns on specific aspects of geographic information, such as scale (levels of detail, accuracy, uncertainty) and the need to integrate data from various sources. Geographic information from each source needs to be consolidated in order to be valuable to high-level decision-makers. In this case, SDI can be seen as a set of building blocks, in which hierarchies are built through the exchange and consolidation of information from corporate and local levels, to regional and global levels. In this hierarchy, lower levels (Davis & Alves, 2005) provide detailed information that helps to consolidate the upper, more general, levels (Rajabifard & Williamson, 2001). The integration problem also requires attention to semantics, because data produced by different organizations, for different needs, are not necessarily compatible, even if they refer to the same location or to the same real-world subject. In this particular issue, the development and use of ontologies may be required.

Finally, there is a technological level. The exchange of information can occur in several ways, but the most interesting one is the use of Web services, using a service-based architecture approach, thus achieving *loosely-coupled and distributed geographic information systems* (Bernard & Craglia, 2005; Davis & Alves, 2005). There are pending issues related to the compatibility between Web service standards defined by the OGC and by the World Wide Web Consortium (W3C), but there are already initiatives to bridge them (Bacharach, 2007; Kim, Kim, Lee, & Joo, 2005). There is also the need to define and propose higher-level services, so SDI can go beyond the simple discovery and download of geographic data, and provide solutions to location-related problems using multiple and distributed sources of information.

BACKGROUND

Creating geographic datasets is a complex and expensive undertaking. In the past, redundant efforts in dataset creation were commonplace: organizations with an interest in the same areas, therefore potential partners for sharing basic data, would not cooperate due to their diverse technological strategies, budgeting, and timing. Of course, such redundancy was undesirable, motivating the creation of cooperation efforts for data sharing.

An example of such an arrangement took place in Belo Horizonte, Brazil. A cooperation agreement, involving 29 different organizations, including government agencies, universities, and private-sector companies, has been active since 1994 (Davis & Fonseca, 2005). Even though political and organizational problems have been solved in this case, technologically there is still a lot to be done. Data for interchange reside in a FTP server, for download by authorized people in one of five different data formats. There is also a metadata sheet for each information class, presented as a simple and nonstandardized text file. Simple as it may seem, maintaining such a setting requires much effort by those who coordinate data gathering and distribution, because much work is performed manually or with little automation. Constant and efficient interpersonal communication is required, so that one organization's needs on some data can be informed to the organizations that generate that data. Thus, cooperative settings are a great improvement on the early approach to sharing, but their scalability potential is rather limited.

Data Transfer Standards

Large-scale off-line data sharing depends fundamentally on data translation, because each organization can use a different geographic information system (GIS). Many efforts in the past have tried to establish a neutral file format for exchange purposes, so that every GIS would only need translators to and from this common format (Lima, Câmara, & Monteiro, 2001).

In practice, commercial formats are used in most data transfer situations, reflecting the influence of the user community of a given GIS package. Regardless of using *de facto* or *de jure* standards, this approach addresses syntactic concerns only, avoiding semantic issues. Furthermore, data transfer formats are unsuitable for online access, maintaining the need for an export-import off-line cycle. Off-line sharing causes multiple copies of the same data to be distributed among interested parties at different times, causing serious synchronization problems.

Spatial Data Clearinghouses

From the establishment of a standard (or, at least, from some *de facto* standards), many national mapping agencies started to create *spatial data clearinghouses*, Internet-based settings that intend to facilitate access to spatial data. A centralized site, from which data from several sources can be found, is established, including services for searching, viewing, transferring, and ordering spatial data (Crompvoets, Bregt, Rajabifard, & Williamson, 2004). Clearinghouses allow data providers to make their offerings known by users, with descriptions (metadata) and instructions on how to access and use the data.

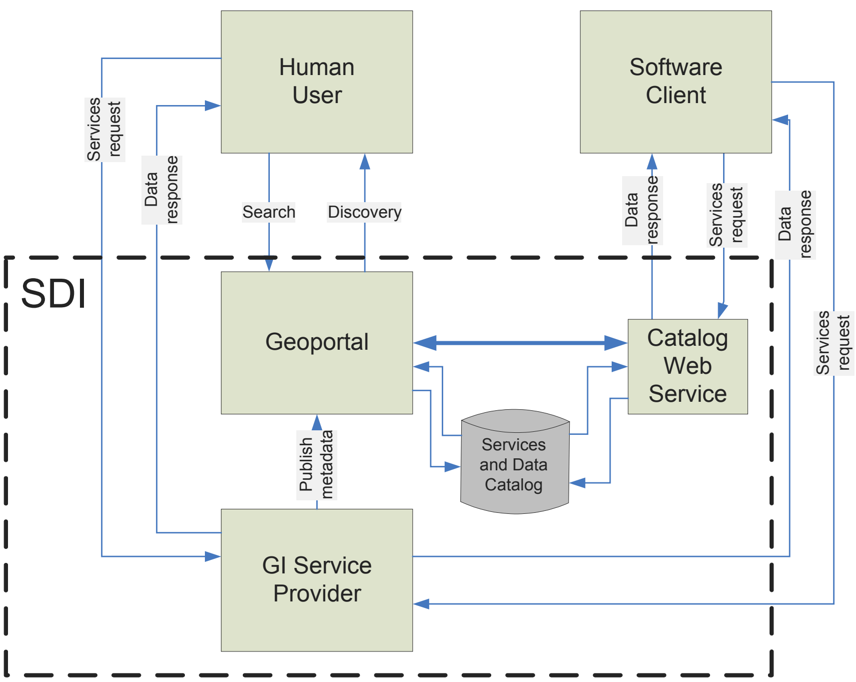
Clearinghouses have been more recently described as a kind of Web portal, that is, a site or a gateway through which commonly used services are offered (INSPIRE, 2002). The emphasis on services is recent, compared to previous implementations, which were mostly based on a combination of technical tools, institutional cooperation mechanisms, and commercial concerns, directed at "off-the-shelf" data dissemination (FGDC, 1997).

A recent study on clearinghouses (Crompvoets et al., 2004) showed that users are dissatisfied with their functionality, indicating that the focus should change from a data-oriented to a user- and application-oriented view. This can be achieved by using service-based architectures.

Early Spatial Data Infrastructures

The expression "spatial data infrastructure" was initially used to describe the provision of standardized access to geographic information (Maguire & Longley, 2005). Many clearinghouse initiatives evolved to what Masser (1999) calls "the first generation of national spatial data infrastructures," while observing that "infrastructure" implies the existence of some sort of coordination for policy formulation and implementa-

Figure 1. Geoportals and SDI



Related Content