Handbook of Research on Geoinformatics

Hassan A. Karimi
University of Pittsburgh, USA

INFORMATION SCIENCE REFERENCE
Hershey · New York

Director of Editorial Content: Kristin Klinger
Director of Production: Jennifer Neidig
Managing Editor: Jamie Snavely
Assistant Managing Editor: Carole Coulson
Typesetter: Jeff Ash
Cover Design: Lisa Tosheff
Printed at: Yurchak Printing Inc.

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue, Suite 200
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

and in the United Kingdom by
Information Science Reference (an imprint of IGI Global)
3 Henrietta Street
Covent Garden
London WC2E 8LU
Tel: 44 20 7240 0856
Fax: 44 20 7379 0609
Web site: http://www.eurospanbookstore.com

Library of Congress Cataloging-in-Publication Data

Handbook of research on geoinformatics / Hassan A. Karimi, editor.

p. cm.

Includes bibliographical references and index.

Summary: "This book discusses the complete range of contemporary research topics such as computer modeling, geometry, geoprocessing, and geographic information systems"--Provided by publisher.

ISBN 978-1-59904-995-3 (hardcover) -- ISBN 978-1-59140-996-0 (ebook)

1. Geographic information systems--Research--Handbooks, manuals, etc. I. Karimi, Hassan A.

G70.212.H356 2009

910.285--dc22

2008030767

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter V
Spatial Data Infrastructures

Carlos Granell
Universitat Jaume I, Spain

Michael Gould
Universitat Jaume I, Spain

Miguel Ángel Manso
Technical University of Madrid, Spain

Miguel Ángel Bernabé
Technical University of Madrid, Spain

ABSTRACT

Geographic Information Systems (GIS) are data-centric applications that rely on the input and constant maintenance of large quantities of basic and thematic spatial data in order to be useful tools for decision-making. This chapter presents the institutional collaboration framework and the major technology components to facilitate discovery and sharing of spatial data: Spatial Data Infrastructures (SDI). We review the essential software components –metadata editors and associated catalogue services, spatial data content repositories, client applications, and middleware or intermediate geospatial services– that define SDIs as heterogeneous distributed information systems. Finally we highlight future research needs in the areas of semantic interoperability of SDI services and in improved institutional collaboration.

INTRODUCTION

Geographic Information Systems (GIS) and related spatial applications are data-centric in the sense that they rely on the input and constant maintenance of large quantities of reference spatial data, on top of which integrators and end-users produce value-added thematic geographic information for the purpose of decision-making. A typical GIS workflow can be simplified as consisting of three components: 1) data entry and reformatting, 2) data processing (geoprocessing), and 3) presentation of results to the user. In practice, this apparently simple workflow is constrained by two key factors. The first is limited interoperability among GIS components, because most are tightly coupled to specific data formats or to other software, complicating the task of integrating components from multiple vendors. The second is that the basic spatial data (reference data) necessary to begin geoprocessing are in many cases not readily available, because they are poorly documented, outdated, are too expensive, or are available under restrictive licensing conditions. This second factor has been seriously limiting the ability of government employees, researchers and businesses to exploit geographic information, unnecessarily incrementing project costs and, thus, negatively affecting the economy.

Many government administrations have recognized this critical problem and have initiated coordinated actions to facilitate the discovery and sharing of spatial data, creating the institutional basis for Spatial Data Infrastructures (SDI) (van Loenen and Kok 2004). The Global Spatial Data Infrastructure (GSDI) association (www.gsdi.org) defines SDI as a *coordinated series of agreements on technology standards, institutional arrangements, and policies that enable the discovery and facilitate the availability of and access to spatial data*. The SDI, once agreed upon and implemented, serves to connect GIS and other spatial data users to a myriad of spatial data sources, the majority of which are held by public sector agencies.

In 1990 the U.S. Federal Geographic Data Committee (FGDC) was created and in 1994, then president, William Clinton, asked it (Executive Order 12906) to establish a national SDI in conjunction with organizations from state, local, and tribal governments, the academic community, and the private sector. Three years later the European Umbrella Organization for Geographic Information (EUROGI) was created with the mission to develop a unified European approach to the use of geographic technologies (a mission far from complete). More recently, the European Commission launched the Infrastructure for Spatial Information in Europe (INSPIRE) initiative for the creation of a European Spatial Data Infrastructure (ESDI), based on a Framework Directive (European legislation) defining how European member states should go about facilitating discovery and access to integrated and interoperable spatial information services and their respective data sources. As the number of national SDIs increased, to include in 2004 about half the nations worldwide (Masser 2005; Crompvoets et al. 2004), the Global Spatial Data Infrastructure (GSDI) Association was created to promote international cooperation and collaboration in support of local, national, and international SDI developments.

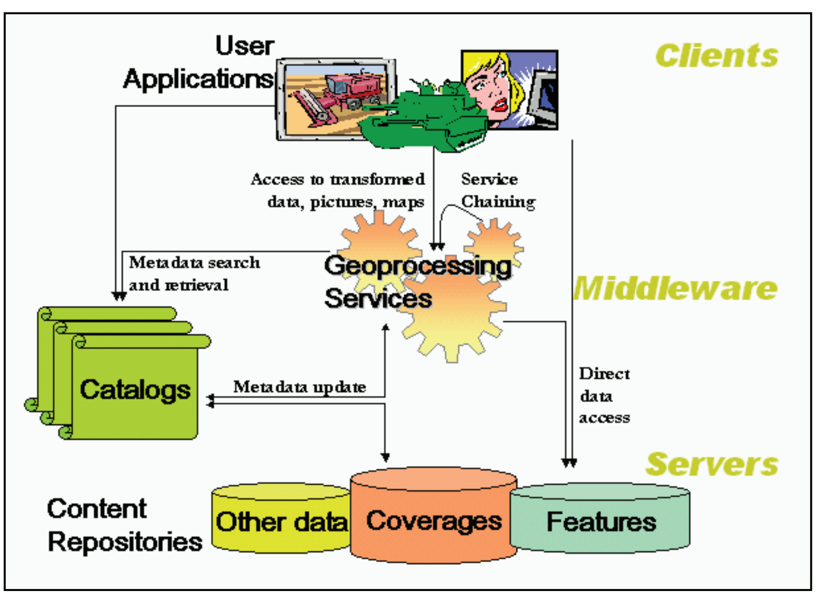
The basic creation and management principles of SDI apply to any and all spatial jurisdictions in a spatial hierarchy, from municipalities to regions, states, nations, and international areas. Béjar et al. (2004) show how each SDI at each level in the hierarchy can be created in accordance with its thematic (e.g., soils, transportation) and geographical (e.g., municipality, nation) coverage, following international standards-based processes and interfaces, to help ensure that the SDIs fit like puzzle pieces, both geographically and vertically (thematically). This harmonization exercise is necessary to allow for seamless spatial data discovery and exploitation crossing jurisdictional boundaries, in the case of response to flooding or forest fires, just to name two important cross-border applications. In practice this harmonization has been difficult to achieve due to political but also semantic-related differences between neighboring regions. An early (1980s) European exercise in cross-border harmonization, stitching together nationally-produced pieces of the Coordinated Information on the European Environment (CORINE) land cover database, highlighted some of these discrepancies at regional and national borders: experts on both sides disagreed on how to classify the same, cross-border land cover regions.

SDI ESSENTIAL COMPONENTS

Although SDIs are primarily institutional collaboration frameworks, they also define and guide implementation of heterogeneous distributed information systems, consisting of four main software components linked via Internet. These components are: 1) metadata editors and associated catalogue services, 2) spatial data content repositories, 3) client applications for user search and access to spatial data, and 4) middleware or intermediate geoprocessing services which assist the user in finding and in transforming spatial data for use at the client side application.

Figure 1 summarizes these essential technology components, as generally accepted within the geographic information standards organizations Open Geospatial Consortium (OGC) (www.opengeospatial.org) and ISO Technical Committee 211 (www.isotc211.org), and synthesized by the FGDC and NASA. This conceptual architecture may be interpreted as a traditional 3-tier client-middleware-server model, where GI applications seek spatial data content that are discovered and then possibly transformed or processed by intermediary services before presentation by the client application. But the architecture also may be interpreted using the web services 'publish-find-bind' triangle model (Gottschalk et al. 2002), whereby spatial data content (and service) offers are published to catalogue servers, which are later queried to discover (find) data or services, and then the client application binds to (consumes or executes) them.

Figure 1. High-level SDI architecture, taken from the FGDC-NASA Geospatial Interoperability Reference Model (GIRM), (FGDC 2003)



Regardless of the precise conceptual model adopted, what is common among nearly all SDIs is the primary goal of improving discovery and access to spatial data. Discovery is based on the documentation of datasets to be shared, in the sense of metadata following international standards such as ISO 19115/19139. Metadata describing the content, geographic and temporal coverage, authorship, access and usage rights details, and other attributes of a dataset are created within GIS applications or externally using specialized text editors. The metadata files are stored in standard XML formats and are then sent (published) to some data catalogue server, in many cases one which is located at a central node of the SDI but in principle may be distributed anywhere on the network.

Users wishing to discover spatial data sources normally access catalogue search interfaces via web applications called Geoportals (Bernard et al. 2005), examples of which may be found at http://www.geo-one-stop.gov/ and http://eu-geoportal.jrc.it/. The geoportal is an interface façade, both hiding the implementation details of the underlying catalogue query mechanisms, and inviting participation in the SDI community. In addition to discovery queries, the geoportal also normally provides free access to quick looks or small samples of datasets that are discovered. This spatial data visualization is frequently implemented as software employing Web Map Service (WMS) software interfaces (OGC 2006), allowing for integration of heterogeneous client and server products from multiple vendors, as proprietary or free software solutions. WMS-based services receive a request for a certain spatial data layer and for a certain geographical extent, convert the data (initially in vector or raster format) to create a bitmap (standard MIME formats such as JPEG, GIF, PNG) and then deliver the image to the web client (browser or GIS/SDI client).

More sophisticated spatial data (web) services are becoming available, many of which also following *de jure* ISO standards and *de facto* specifications from organizations such as OGC, Organization for the Advancement of Structured Information Standards (OASIS), and W3C. These include services providing concrete functionality such as coordinate transformation, basic image processing and treatment, basic geostatistics, and composition or chaining of individual services to form more complex services.

Summarizing both the institutional and technological aspects, article 1 of the INSPIRE European Directive proposal (EC 2004), lists the following necessary components for SDIs:

"The component elements of those infrastructures shall include metadata, spatial datasets and spatial data services; network services and technologies; agreements on sharing, access and use; and coordination and monitoring mechanisms, processes and procedures."

Caution should be taken when describing more narrow initiatives, projects or products which provide only a subset of these requirements, as SDI. This is especially the case of institutional agreements and cooperation for improved access to spatial data: the above principles and components should be explicitly involved.

FUTURE RESEARCH

SDI researchers are active on several fronts, but two main areas of interest are: improving institutional collaboration and SDI effectiveness (including cost-benefit analyses and more elaborate data access policy), and SDI component implementation and testing. In the second area fall topics such as semantic interoperability and composition of SDI (web) services, the integration of so-called disruptive technologies such as Google Earth and similar commercial services, grass-roots initiatives contributing user-generated data, integration with grid computing and with e-Government solutions, and exploitation of data from diverse sensor networks.

For further details on SDI the reader should consult the Spatial Data Infrastructure Cookbook (GSDI 2004) and the European Commission's International Journal of SDI Research (http://ijsdir.jrc.it).

REFERENCES

Béjar, R., Gallardo, P., Gould, M., Muro, P., Nogueras, J., & Zarazaga, J. (2004). A high level architecture for national SDI: The Spanish case. *EC-GI&GIS Workshop*, Warsaw, June 2004. Retrieved April 4, 2006, from http://www.ec-gis.org/Workshops/10ec-gis/.

Bernard, L., Kanellopoulos, I., Annoni, A., & Smits, P. (2005). The European geoportal — one step towards the establishment of a European Spatial Data Infrastructure. *Computers, Environment and Urban Systems*, *29*(1), 15–31.

Crompvoets, J., Bregt, A., Rajabifard, A., & Williamson, I. (2004). Assessing the worldwide status of national spatial data clearinghouses. *International Journal of Geographical Information Science*, *18*(7), 665-689.

EC -- Commission of The European Communities (2004). Proposal for a Directive of the European Parliament and of the Council establishing an infrastructure for spatial information in the Community (INSPIRE), COM(2004) 516. Retrieved April 4, 2006, from http://inspire.jrc.it/proposal/EN.pdf.

FGDC (2003). The Geospatial Interoperability Reference Model, version 1.1. Federal Geographic Data Committee Geospatial Applications Interoperability (GAI) Working Group. Retrieved April 4, 2006, http://gai.fgdc.gov/girm/v1.1/.

Gottschalk, K., Graham, S., Krueger, S., & Snell, J. (2002). Introduction to Web services architecture. *IBM Systems Journal, 41*(2). Retrieved April 4, 2006, from http://researchweb.watson.ibm.com/journal/sj/412/gottschalk.html

GSDI (2004). Spatial Data Infrastructure Cookbook (English version 2.0). Retrieved April 4, 2006, from http://www.gsdi.org/gsdicookbookindex.asp.

Masser, I. (2005). *GIS Worlds; Creating Spatial Data Infrastructures*. Redlands, California: ESRI Press.

OGC (2006). OpenGIS Web Map Service (WMS) implementation specification, version 1.3. Retrieved April 4, 2006, from http://portal.opengeospatial.org/files/?artifact_id=14416.

van Loenen, B., & Kok, B.C. (Eds.). (2004). *Spatial data infrastructure and policy development in Europe and the United States*. Delft: Delft University Press.

KEY TERMS

FGDC: Federal Geographic Data Committee, an interagency committee established in the US in 1990, with the mandate to create and support data sharing, in the form of the US National SDI. http://www.fgdc.gov.

GI: Geographic information, the subset of information pertaining to, or referenced to, known locations on or near the Earth's surface.

GSDI: Global Spatial Data Infrastructure Association. An umbrella organization grouping national, regional and local organizations dedicated to the creation and maintenance of SDIs around the world. http://www.gsdi.org.

OGC: Open Geospatial Consortium, a membership body of 300-plus organizations from the commercial, government and academic sectors, that creates consensus interface specifications in an effort to maximize interoperability among software detailing with geographic data. http://www.opengeospatial.org.

SDI: Spatial Data Infrastructure

WMS: Web Map Service, a software interface specification published by the Open Geospatial Consortium (OGC). The specification defines how software clients should formulate queries to compliant map servers, and how those servers should behave and respond.