

The Handbook of Geographic Information Science

Edited by

John P. Wilson

and

A. Stewart Fotheringham

The Handbook of Geographic Information Science

Blackwell Companions to Geography

Blackwell Companions to Geography is a blue-chip, comprehensive series covering each major subdiscipline of human geography in detail. Edited and contributed to by the disciplines' leading authorities each book provides the most up to date and authoritative syntheses available in its field. The overviews provided in each Companion will be an indispensable introduction to the field for students of all levels, while the cutting-edge, critical direction will engage students, teachers, and practitioners alike.

Published

1. *A Companion to the City*

Edited by Gary Bridge and Sophie Watson

2. *A Companion to Economic Geography*

Edited by Eric Sheppard and Trevor J. Barnes

3. *A Companion to Political Geography*

Edited by John Agnew, Katharyne Mitchell, and Gerard Toal
(Gearoid O Tuathail)

4. *A Companion to Cultural Geography*

Edited by James S. Duncan, Nuala C. Johnson, and Richard H. Schein

5. *A Companion to Tourism*

Edited by Alan A. Lew, C. Michael Hall, and Allan M. Williams

6. *A Companion to Feminist Geography*

Edited by Lise Nelson and Joni Seager

7. *The Handbook of Geographic Information Science*

Edited by John P. Wilson and A. Stewart Fotheringham

The Handbook of Geographic Information Science

Edited by

John P. Wilson

and

A. Stewart Fotheringham



Blackwell
Publishing

© 2008 by Blackwell Publishing Ltd

BLACKWELL PUBLISHING

350 Main Street, Malden, MA 02148-5020, USA

9600 Garsington Road, Oxford OX4 2DQ, UK

550 Swanston Street, Carlton, Victoria 3053, Australia

The right of John P. Wilson and A. Stewart Fotheringham to be identified as the Authors of the Editorial Material in this Work has been asserted in accordance with the UK Copyright, Designs, and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs, and Patents Act 1988, without the prior permission of the publisher.

First published 2008 by Blackwell Publishing Ltd

1 2008

Library of Congress Cataloging-in-Publication Data

The handbook of geographic information science / edited by John P. Wilson and A. Stewart Fotheringham.

p. cm. — (Blackwell companions to geography)

Includes bibliographical references and index.

ISBN 978-1-4051-0795-2 (hardback : alk. paper) — ISBN 978-1-4051-0796-9 (pbk. : alk. paper)

1. Geography—Data processing. 2. Geographic information systems. I. Wilson, John P. (John Peter), 1955– II. Fotheringham, A. Stewart.

G70.2.H356 2008

910.285—dc22

2007008297

A catalogue record for this title is available from the British Library.

Set in 10/12pt Sabon

by Graphicraft Limited, Hong Kong

Printed and bound in Singapore

by Utopia Press Pte Ltd

The publisher's policy is to use permanent paper from mills that operate a sustainable forestry policy, and which has been manufactured from pulp processed using acid-free and elementary chlorine-free practices. Furthermore, the publisher ensures that the text paper and cover board used have met acceptable environmental accreditation standards.

For further information on

Blackwell Publishing, visit our website:

www.blackwellpublishing.com

Contents

<i>List of Figures</i>	viii
<i>List of Contributors</i>	xiv
Geographic Information Science: An Introduction	1
Part I Data Issues	9
1 The Availability of Geographic Data: The Current Technical and Institutional Environment <i>David J. Cowen</i>	11
2 Social Data <i>David J. Martin</i>	35
3 Remote Sensing <i>Brian G. Lees</i>	49
4 Spatialization <i>André Skupin and Sara I. Fabrikant</i>	61
5 Uncertainty in Spatial Databases <i>Ashley Morris</i>	80
6 On the Identification of Uncertainties in Spatial Data and Their Quantification with Probability Distribution Functions <i>James D. Brown and Gerald B. M. Heuvelink</i>	94
Part II Database Trends and Challenges	109
7 Object-Oriented Database Management Systems <i>Shashi Shekhar and Ranga Raju Vatsavai</i>	111
8 Adding the Z Dimension <i>Michael F. Hutchinson</i>	144

9	Adding Time into Geographic Information System Databases <i>May Yuan</i>	169
10	Geospatial Data Integration <i>Craig A. Knoblock and Cyrus Shahabi</i>	185
Part III	Visualization	197
11	Mapping in a Digital Age <i>William E. Cartwright</i>	199
12	Generalization of Spatial Databases <i>William A. Mackaness</i>	222
13	Geographic Information Systems and Surfaces <i>Nicholas J. Tate, Peter F. Fisher, and David J. Martin</i>	239
14	Fuzzy Classification and Mapping <i>Vincent B. Robinson</i>	259
15	Rule-Based Mapping <i>A-Xing Zhu</i>	273
16	Multivariate Geovisualization <i>Mark Gahegan</i>	292
17	Virtual Reality in Geographic Information Systems <i>Michael Batty</i>	317
Part IV	Knowledge Elicitation	335
18	Inference and Spatial Data <i>Chris Brunsdon</i>	337
19	Geographic Data Mining and Knowledge Discovery <i>Harvey J. Miller</i>	352
20	The Geospatial Semantic Web <i>Frederico Fonseca</i>	367
Part V	Spatial Analysis	377
21	Quantitative Methods and Geographic Information Systems <i>Martin E. Charlton</i>	379
22	Spatial Cluster Analysis <i>Geoffrey M. Jacques</i>	395
23	Terrain Analysis <i>Yongxin Deng, John P. Wilson, and John C. Gallant</i>	417
24	Dynamic Modeling <i>Jochen Albrecht</i>	436

Part VI	Geographic Information Systems and Society	447
25	Institutional Geographic Information Systems and Geographic Information Partnering <i>David L. Tulloch</i>	449
26	Participatory Geographic Information Systems <i>Daniel Weiner and Trevor M. Harris</i>	466
27	Geographic Information Systems and Participatory Decision Making <i>Piotr Jankowski and Timothy L. Nyerges</i>	481
28	Surveys of People and Place <i>Peter H. Dana</i>	494
29	Geographic Information Science, Personal Privacy, and the Law <i>George C. H. Cho</i>	519
30	Geographic Information Systems in Education <i>Joseph J. Kerski</i>	540
Part VII	Future Trends and Challenges	557
31	Web-based Geographic Information Systems <i>Christopher B. Jones and Ross S. Purves</i>	559
32	Location-based Services and Geographic Information Systems <i>Allan J. Brimicombe</i>	581
33	Geographic Information Science: The Grand Challenges <i>Michael F. Goodchild</i>	596
34	Geographic Information Science: Where Next? <i>Andreas Reuter and Alexander Zipf</i>	609
	<i>Index</i>	620

Figures

0.1	Geographic Information Science (GISc): An Overview	2
1.1	Geographic data marketplace	13
1.2	Data flow to and from government	16
1.3	Fundamental issues in spatial data transfer	18
1.4	The Spatial Data Transfer Standard	18
1.5	Content standard for digital geospatial metadata	22
1.6	FGDC decision tree	23
1.7	FGDC steps to create a clearinghouse	24
1.8	Geospatial One-Stop	26
1.9	US Fish and Wildlife Service Wetlands Mapper	27
1.10	The National Map	28
1.11	The National Map Seamless Data Distribution System Viewer	29
1.12	The National Atlas	31
1.13	The Geography Network	32
2.1	Comparison of principal social data sources	38
3.1	Some typical platforms and instrument packages	50
4.1	Census-based visualization of trajectories of Texas counties	65
4.2	Cases of convergence and divergence in a spatialization of Texas county trajectories	65
4.3	Scatter plots derived from demographic data for US states	66
4.4	Spatializations derived from 32 demographic variables	66
4.5	Portion of a spatialization of conference abstracts	69
4.6	Procedure for deriving a spatialization from AAG conference abstracts	70
4.7	Conference abstract as unstructured text	71
4.8	Conference abstract in semi-structured form as part of an XML file	71
4.9	Application of spring model layout and pathfinder network scaling	73
4.10	The tree map method	74
4.11	Use of GIS in implementing scale-dependent spatialization	74

5.1	Case in which the points are stored in the database and connected in order from the last point back to the first point	84
5.2	Case in which the segments between the points are connected in no particular order	84
7.1	Evolution of databases	112
7.2	Census blocks with boundary ID:1050	115
7.3	Four tables required in a relational database with overlapping attributes	115
7.4	Three-layer architecture	118
7.5	An OGIS proposal for building blocks of spatial geometry in UML notation	119
7.6	The buffer of a river and points within and outside the resultant buffer	135
8.1	Contour, point elevation, and streamline data	148
8.2	Minimum curvature gridding of point elevation data with spurious sinks or depressions	153
8.3	Spurious sinks removed from the surface in Figure 8.2 by the drainage enforcement algorithm	154
8.4	Contours and inferred stream lines and ridge lines derived by the ANUDEM procedure	155
8.5	Plot of root mean square slope of a DEM versus DEM resolution	156
8.6	Spline model of annual mean precipitation over topography	161
9.1	Examples of representations of temporal information in a relational data model	174
9.2	Examples of representations of spatio-temporal information in a GIS environment	175
9.3	A conceptual framework for a three-domain model	177
10.1	A wrapper converts the New York State Property Database into structured data that can then be integrated with other sources	187
10.2	Automatic conflation of vector data with imagery	189
10.3	Integrating and reasoning about the property tax data, satellite imagery, and road vector data to identify the structures in an image	190
10.4	Automatic conflation of maps with imagery	192
10.5	Results of conflating MapQuest map with imagery	192
10.6	Integration of train schedules with vector data and maps	194
11.1	CEEFAQ weather map	202
11.2	Minitel "kiosk"	202
11.3	<i>Aspen Movie Map</i> Project	204
11.4	Surrogate walk from the Domesday Community disc	205
11.5	Territorial evolution of Canada	206
11.6	Ashdowne's <i>Virtual Atlas</i> – initial text interface	207
11.7	Perry-Castañeda Library, The University of Texas at Austin map of Australia	208
11.8	CIA World Fact Book map of Australia	209
11.9	MapQuest – Victoria, Australia	210
11.10	Bodleian Library: Plan for rebuilding . . . London – J Evelyn 1666	211

11.11	LandChannel, Government of Victoria	211
11.12	WhereIs map information	212
11.13	Australian Coastal Atlas	214
11.14	National Atlas of Canada Quick Maps	215
11.15	Atlas of Switzerland	216
11.16	WAP-enabled cellular telephone delivering map information from Webraska	217
12.1	1:25,000, 1:100,000, and 1:250,000	224
12.2	Generalization as a sequence of modeling operations	225
12.3	Model and cartographic generalization acting in unison to reveal different qualities about The Tower of London	225
12.4	Generalization in the context of automated solutions	227
12.5	The choice, sequence, and degree of application of various methods enable synthesis of different solutions	229
12.6	Transformations with decreasing map scale, and corresponding scale bands for a topographic map	231
12.7	Examples drawn from paper maps of building generalization at various scales	231
12.8	Decision tree for key buildings	232
12.9	Different products according to theme and scale derived from the same source	233
13.1	Sample frame for soil fertility characteristics	242
13.2	Sample frame for LiDAR data	243
13.3	A TIN generated from the points displayed in Figure 13.2	245
13.4	LiDAR derived raster hillshade, TIN, and TIN hillshade	246
13.5	Perspective surface views and TINS with and without adaptive LOD proportional to the viewpoint	247
13.6	Kernel estimation of a reference-interval function	248
13.7	An isoline visualization or contour map	249
13.8	A filled contour visualization of the same area as is shown in Figure 13.7	250
13.9	A psuedo-3D view of the same general area as is shown in Figure 13.7	251
13.10	A slope map of the same area as is shown in Figure 13.7	252
13.11	Hillshaded maps of the same area as is shown in Figure 13.7	253
13.12	A navigable 3D view of the area around the University of Leicester	254
14.1	A simple example of typical results from using the FCM algorithm	265
15.1	Knowledge-based approach to rule-based mapping under GIS	275
15.2	Example of descriptive knowledge expressed as a rule for rule-based mapping	275
15.3	A distribution of Soil Series Basco in Pleasant Valley, Wisconsin, USA	280
15.4	The metrics of a fuzzy membership function	281
15.5	Three basic forms of membership functions	282

15.6	The predicted distribution of Soil Series Basco in Pleasant Valley, Wisconsin, USA based on a fuzzy implementation of the rule in Figure 15.2	285
15.7	Predicted distribution of soil series in Pleasant Valley, Wisconsin, USA based on a Boolean implementation of the rules in Figure 15.8	287
15.8	Descriptive rules for soils in Pleasant Valley, Wisconsin, USA	287
15.9	Predicted distribution of soil series in Pleasant Valley, Wisconsin, USA based on a fuzzy implementation of the rules in Figure 15.8	288
15.10	Uncertainty map associated with the soil map produced under the fuzzy implementation of the rules for the Pleasant Valley area	289
16.1	Thematic map of population by county for the conterminous 48 states of the USA	297
16.2	Visualization of changes in population growth and immigrant population using proportional circles	298
16.3	Which inner circle seems bigger? A visual illusion concerning the size of circles	298
16.4	Characterization of oil reservoirs through time	299
16.5	A parallel coordinate plot (PCP) showing demographic and related information for the 48 conterminous US states	300
16.6	Multiple scatterplots and bivariate maps of demographics and cancer-related data, organized in a matrix	301
16.7	Scatterplot matrix showing land cover data and decision tree rules used in classifying the data	302
16.8	A surface produced by clustering US states based on their demographic properties	303
16.9	The city of London during the seventeenth century	304
16.10	A concept map depicting events that have shaped the landscape of central Pennsylvania	305
16.11	A visualization of 100 census variables showing correlation and entropy	306
17.1	The basic geometric model built in <i>ArcScene</i> and an example of a finely rendered building imported into the scene	321
17.2	A geographic “sea” showing the intensity of retail activity layered onto the 3D geometry of central London	323
17.3	A typical fly-through in virtual London	324
17.4	Porting the digital model and various panoramas into a virtual world	326
17.5	The Arthur interface	327
17.6	Geographic visualization of web traffic	329
17.7	Interfaces to GIS using 3D globes	330
17.8	Constructing the Simulacra	332
17.9	Printing the virtual city using CAD/CAM technology	333
18.1	The location of study area (LHS) and the houses in the samples (RHS)	342
18.2	Posterior distribution for $\delta = \mu_1 - \mu_2$	345

19.1	A concept hierarchy for <i>location</i>	355
19.2	A data cube for traffic data attributed by date, time of day, and station	356
19.3	Traffic map cube visualization by date, time of day, and station	360
19.4	Traffic map cube visualization by date, time of day, and geographic location	361
22.1	Cluster morphology of cancer incidence on Long Island	405
22.2	Cluster change and persistence	406
22.3	Cluster change and persistence continued	407
22.4	Cluster change and persistence continued	408
22.5	Disparity clusters	410
22.6	Information frames provide cogent summaries of the properties of spatial cluster statistics	412
23.1	Schematic diagram showing site-specific, local, and regional interaction as a function of time	419
23.2	Watershed modeling with GIS and WEPP	424
23.3	Scale effects of terrain analysis	430
24.1	Eight possible ATC commands for a single aircraft	440
25.1	A conceptual diagram of multipurpose land information system development	453
26.1	The relations of GIS, participation, and the Internet in the case studies	470
28.1	CCARC Central American mapping projects	496
28.2	CCARC Nicaraguan map detail with large point symbols and line hatching indicating survey imprecision	499
28.3	Tiling and overlaps in CCARC Nicaraguan <i>Bloques</i>	500
28.4	Draft map of Krukira with straight lines between points	501
28.5	Sketch map of Auka, Honduras	502
28.6	Topographic map of Auka, Honduras	504
28.7	Waterproof kit contents	505
28.8	DGPS receiver and pocket PC with ArcPad data collection software	510
28.9	GPS waterproof notebook pages	511
28.10	Position types: Direct, indirect, registered, and estimated	512
28.11	Land use categories and symbols for Nicaragua	513
28.12	Land use categories and symbols for Honduras	514
28.13	Final map of Auka, Honduras	515
28.14	Land use differences in overlap between Ahuas and Wawina, Honduras	516
30.1	PC monitor in Texas high school showing series of themes for Africa	541
30.2	Dimensions of GIS education	542
30.3	Student displaying the results of his GIS analysis in 3D	552
31.1	Example from the Pont map collection	561
31.2	City of Vallejo economic development information system	563
31.3	Options for thematic map and resulting presentation of Scottish census data	564

31.4	The interface of the USGS seamless data distribution system	566
31.5	An example search for data with GeoXWalk	567
31.6	Three-tier web GIS architecture	570
31.7	Open web services framework	572
31.8	Interoperating web map Services	573
31.9	Two representations of the same data using different style layer descriptors and an extract from the SLD	574
31.10	A thin client and a thick client	576
31.11	Ordnance Survey MasterMap data displayed as SVG	576
31.12	Example of the use of SVG	577
32.1	The technological niche of LBS	583
32.2	Experimental PDA interface for LBS that provides map, text and voice wayfinding instructions	585
32.3	Basic system architecture for LBS	586
34.1	Object-centric location models	616

Contributors

Jochen Albrecht

Department of Geography, Hunter College, City University of
New York, 695 Park Avenue, New York, NY 10021, USA.
E-mail: jochen@geo.hunter.cuny.edu

Michael Batty

Centre for Advanced Spatial Analysis, University College London, 1–19
Torrington Place, London WC1E 6BT, UK. E-mail: mbatty@geog.ucl.ac.uk

Allan J. Brimicombe

Centre for Geo-Information Studies, University of East London,
Longbridge Road, Dagenham, Essex RM8 2AS, UK.
E-mail: a.j.brimicombe@uel.ac.uk

James D. Brown

Institute for Biodiversity and Ecosystem Dynamics, Universiteit van
Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands.
E-mail: brown@science.uva.nl

Chris Brunsdon

Department of Geography, University of Leicester, University Road, Leicester
LE1 7RH, UK. E-mail: cb179@lei.ac.uk

William E. Cartwright

School of Mathematical and Geospatial Sciences, RMIT University, Melbourne,
Victoria 3001, Australia. E-mail: w.cartwright@rmit.edu.au

Martin E. Charlton

National Centre for Geocomputation, National University of Ireland, Maynooth,
Ireland. E-mail: martin.charlton@may.ie

George C. H. Cho

Division of Health, Design and Science, University of Canberra, Canberra, ACT 2601, Australia. E-mail: george.cho@canberra.edu.au

David J. Cowen

Department of Geography, University of South Carolina, Columbia, SC 29208, USA. E-mail: cowend@gwm.sc.edu

Peter H. Dana

PO Box 1297, Georgetown, TX 78627, USA. E-mail: pdana@pdana.com

Yongxin Deng

Department of Geography, Western Illinois University, Macomb, IL 61455-1390, USA. E-mail: Y-Deng2@wiu.edu

Sara I. Fabrikant

Department of Geography, University of Zurich, Winterthurestrasse 190, CH-8057 Zurich, Switzerland. E-mail: sara@geo.unizh.ch

Peter F. Fisher

School of Informatics, City University, Northampton Square, London, EC1V 0HB, UK. E-mail: pff1@soi.city.ac.uk

Frederico Fonseca

School of Information Sciences and Technology, Pennsylvania State University, University Park, PA 16802-6823, USA. E-mail: fredfonseca@ist.psu.edu

A. Stewart Fotheringham

National Centre for Geocomputation, National University of Ireland, Maynooth, Ireland. E-mail: stewart.fotheringham@may.ie

Mark Gahegan

GeoVISTA Center, Department of Geography, Pennsylvania State University, University Park, PA 16802, USA. E-mail: mng1@psu.edu

John C. Gallant

CSIRO Land and Water, GPO Box 1666, Canberra ACT 2601, Australia. E-mail: john.gallant@csiro.au

Michael F. Goodchild

National Center for Geographic Information and Analysis, Department of Geography, University of California, Santa Barbara, CA 93106-4060, USA. E-mail: good@geog.ucsb.edu

Trevor M. Harris

Department of Geology and Geography, West Virginia University, Morgantown, WV 26506, USA. E-mail: trevor.harris@mail.wvu.edu

Gerard B.M. Heuvelink

Laboratory of Soil Science and Geology, Wageningen University and Research Centre, PO Box 37, 6700 AA, Wageningen, The Netherlands.
E-mail: gerard.heuvelink@wur.nl

Michael F. Hutchinson

Centre for Resource and Environmental Studies, Australian National University, Canberra ACT 0200, Australia. E-mail: hutch@cres.anu.edu.au

Geoffrey M. Jacquez

BioMedware, Inc., 515 North State Street, Ann Arbor, MI 48104-1236, USA.
E-mail: jacquez@biomedware.com

Piotr Jankowski

Department of Geography, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182-4493, USA. E-mail: piotr@typhoon.sdsu.edu

Christopher B. Jones

Department of Computer Science, Cardiff University, Newport Road, PO Box 916, Cardiff, Wales CF24 3XF, UK. E-mail: c.b.jones@cs.cardiff.ac.uk

Joseph J. Kerski

Environmental Systems Research Institute, Inc., 1 International Court, Broomfield, CO 80021-3200, USA. E-mail: jkerski@esri.com

Craig A. Knoblock

Information Sciences Institute, University of Southern California, Marina del Rey, CA 90292, USA. E-mail: knoblock.isi.edu

Brian G. Lees

School of Physical, Environmental and Mathematical Sciences, University of New South Wales at the Australian Defence Force Academy, Canberra, ACT 2600, Australia. E-mail: b.lees@adfa.edu.au

William A. Mackaness

School of Geosciences, University of Edinburgh, Drummond Street, Edinburgh, Scotland, EH8 9XP, UK. E-mail: william.mackaness@ed.ac.uk

David J. Martin

School of Geography, University of Southampton, Highfield, Southampton SO17 1BJ, UK. E-mail: d.j.martin@soton.ac.uk

Harvey J. Miller

Department of Geography, University of Utah, Salt Lake City, UT 84112-9155, USA. E-mail: harvey.miller@geog.utah.edu

Ashley Morris

School of Computer Science, Telecommunications, and Information Systems, DePaul University, Chicago, IL 60604, USA. E-mail: amorris@cti.depaul.edu

Timothy L. Nyerges

Department of Geography, University of Washington, Seattle, WA 98195-3550, USA. E-mail: nyerges@u.washington.edu

Ross S. Purves

Department of Geography, University of Zurich – Irchel, Winterthurerstr. 190, CH-8057 Zurich, Switzerland. E-mail: rsp@geo.unizh.ch

Andreas Reuter

European Media Laboratory, Schloss-Wolfsbrunnenweg 33, 69118 Heidelberg, Germany. E-mail: andreas.reuter@eml.org

Vincent B. Robinson

Department of Geography, University of Toronto, Mississauga, Ontario L5L 1C6, Canada. E-mail: vbr@eratos.erin.utoronto.ca

Cyrus Shahabi

Integrated Media Systems Center, Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089-2561, USA. E-mail: cshahabi@pollux.usc.edu

Shashi Shekhar

Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455, USA. E-mail: shekhar@cs.umn.edu

André Skupin

Department of Geography, San Diego State University, San Diego, CA 92182-4493, USA. E-mail: skupin@mail.sdsu.edu

Nicholas J. Tate

Department of Geography, University of Leicester, Leicester LE1 7RH, UK. E-mail: n.tate@le.ac.uk

David L. Tulloch

Center for Remote Sensing and Spatial Analysis, Rutgers, State University of New Jersey, New Brunswick, NJ 08904, USA. E-mail: dtulloch@crssa.rutgers.edu

Ranga Raju Vatsavai

Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455, USA. E-mail: vatsavai@cs.umn.edu

Daniel Weiner

Department of Geology and Geography, West Virginia University, Morgantown, WV 26506, USA. E-mail: daniel.weiner@mail.wvu.edu

John P. Wilson

Department of Geography, University of Southern California, Los Angeles, CA 90089-0255, USA. E-mail: jpwilson@usc.edu

May Yuan

Department of Geography, University of Oklahoma, Norman, OK 73019-1007, USA. E-mail: myuan@ou.edu

A-Xing Zhu

Department of Geography, University of Wisconsin, Madison, WI 53706-1404, USA. E-mail: azhu@facstaff.wisc.edu

Alexander Zipf

Department of Geoinformatics and Surveying, University of Applied Sciences FH Mainz, 561156 Mainz, Germany. E-mail: zipf@geoinform.fh-mainz.de

Geographic Information Science: An Introduction

A. Stewart Fotheringham and John P. Wilson

GIS, the acronym for Geographic Information Systems, has been around since the 1980s. Although one can impute a number of characteristics from the use of this acronym, at the heart of the term “systems” lies a computer software package for storing, displaying, and analyzing spatial data. Consequently, the use of the term GIS implies an **object** or tool which one can use for exploring and analyzing data that are recorded for specific locations in geographical space (see Cowen [1988] for an early article articulating this type of definition and Foresman [1998] for a rich and varied account of the history of Geographic Information Systems). Conversely, Geographic Information Science or GI Science, or more simply GISc, represents a much broader framework or *modus operandi* for analyzing spatial data. The term GI Science emphasizes more the **methodology** behind the analysis of spatial data (see Burrough [1986] for what was perhaps the first GIS text to promote such a framework and Chrisman [1999] for an article advocating an extended definition of GIS along these same lines). Indeed, one could define GI Science as: *any aspect of the capture, storage, integration, management, retrieval, display, analysis, and modeling of spatial data*. Synonyms of GI Science include Geocomputation, GeoInformatics, and GeoProcessing.¹

The Breadth of GI Science

Under this definition, GI Science is clearly an extremely broad subject and captures any aspect connected with the process of obtaining information from spatial data. A feeling for this breadth can be seen in Figure 0.1 which describes a schematic of some of the elements that make up GI Science. At the top level, GI Science is concerned with the collection or capture of spatial data by such methods as satellite remotely sensed images, GPS, surveys of people and/or land, Light Detection And Ranging (LiDAR), aerial photographs, and spatially encoded digital video.² The key element here is to capture not only attribute information but also accurate information on the location of each measurement of that attribute. For instance, we might

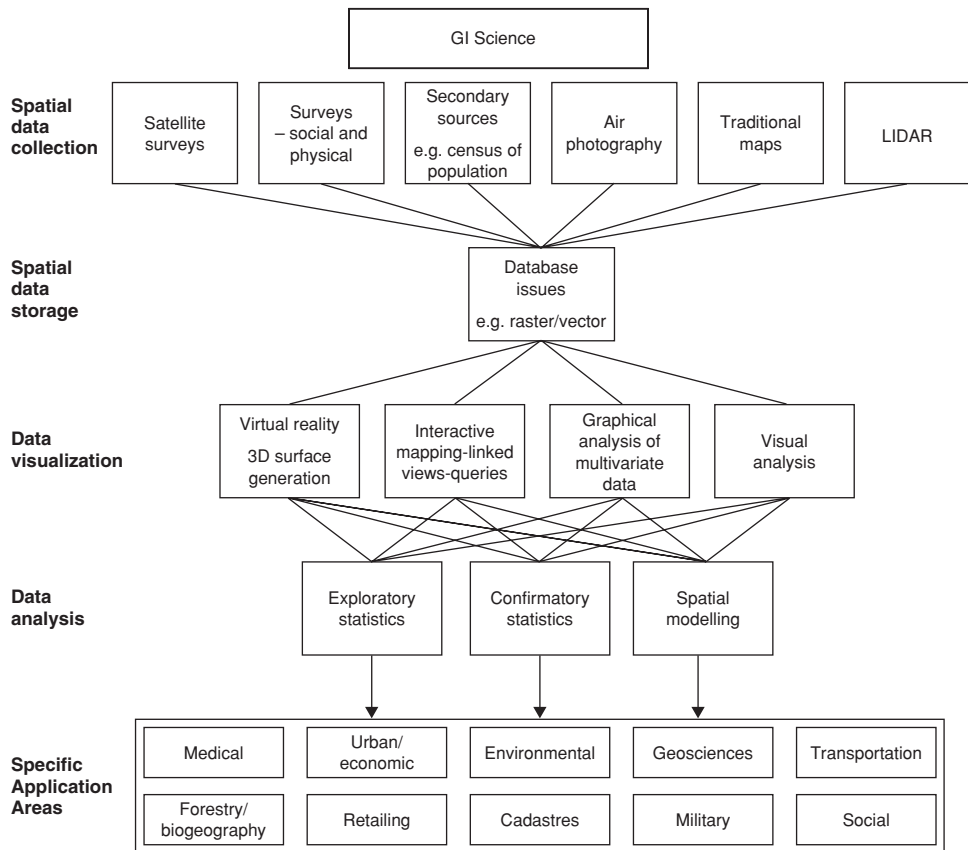


Fig. 0.1 Geographic Information Science (GISc): An Overview

ask people some information on themselves during a survey but we would also like to record some aspect of the location of that individual – this might be the location at which the survey took place, or the person’s usual residence or their workplace or some other location. Similarly, if we measure some attribute such as the elevation above sea level and/or the precipitation at a set of points, we also need to know the locations of these points, otherwise the elevation and precipitation measurements are useless (see Corbett and Carter [1996] and Custer, Farnes, Wilson, and Snyder [1996] for two examples of what can be accomplished by combining locational information with elevations and precipitation measurements).

Once spatial data have been captured, they need to be stored and transmitted. This can create challenges as some spatial data sets can be extremely large. A census of population in the USA would contain, for example, records on almost 300 million people with the locational data on each person typically being his or her residence (hence the well-known problem of census data being a snapshot of where people are at midnight rather than during the day). In some countries, a decennial census has been replaced with a more continuous monitoring of the population in the form of a register which can be updated more regularly. Satellite imagery of the Earth’s

surface can generate terabytes of data and the move towards global data sets can lead to even larger data sets. The data storage and transmission demands of spatially encoded digital video have already been referred to in endnote 2 and pose challenges to current systems. Consequently, spatial data sets can be extremely large and finding ways to store, process, and transmit such large data sets efficiently is a major challenge in GI Science.

The next two levels of operations in the schematic in Figure 0.1 refer to the process of transforming data into information. We are currently living in a data-rich world that is getting richer by the day. In many operations, large volumes of spatial data are being collected and a major challenge in GI Science is to turn these data into useful information. Consider, for example, the following sources of spatial data (which are but a small sample from the complete set of sources):

- Censuses of population which typically occur every five or 10 years and which typically record information on each individual in each household and on the household itself;
- Customer databases held by retail-related companies which hold information on individuals submitted in various application forms or warranty cards;
- Traffic flow monitoring along streets or at intersections;
- LiDAR – low pass fly-overs by plane generating large volumes of detailed data on terrain features or urban areas;
- Digital Elevation Models (DEMs) captured via satellites or the US space shuttle which can be at a global scale;
- Health records either on the location of patients with particular diseases, used to study possible geographic influences on etiology or to assess the level of demand for various services in particular hospitals;
- Satellite remotely-sensed images or aerial photographs used to track land use change over time or to study the spatial impacts of various natural disasters or for various military uses such as tracking missiles or identifying targets;
- Satellite GPS used increasingly for general data capture of vehicles and individuals. This makes possible vehicle tracking, in-car navigation systems, precision agriculture, animal tracking, monitoring of individuals, and general data capture on the location of objects via GPS receivers. It is now possible to contemplate, as the UK is doing, tracking the movement of all vehicles and charging for per mileage road use instead of a flat road tax. Similarly, it is now possible via GPS to monitor a child's movements via a GPS watch linked to a central monitoring system that parents can access remotely via the World Wide Web (see <http://www.wherifywireless.com> for additional details). The linking of GPS to mobile phones will allow the tracking of friends so that one can query the location of a registered friend at any moment. The use of mobile phones to locate individuals by triangulation from mobile relay stations is already standard police practice in the case of missing persons. Some of these uses of course immediately raise important ethical and legal questions which need to be resolved. Just how much spatial information on ourselves are we prepared to have captured and stored?

Most organizations simply do not have the resources (measured in terms of personnel, knowledge, and/or software) to be able to make full use of all the data

they routinely gather. There is a growing need for techniques that allow users to make sense out of their spatial data sets. This mirrors the general transformation of society from one dominated by the industrial revolution with its origins in the eighteenth and nineteenth centuries, to one dominated by the information revolution with its origins in the computer age of the late twentieth century. Two main sets of techniques exist to turn data into information: visualization and statistical/mathematical modeling.

There is a vast array of techniques that have been developed for visualizing spatial data and it remains a very intense and fruitful area of research (see Dykes, MacEachren, and Kraak [2005] for one such treatment). Spatial data lend themselves to visualization because the data are geocoded and can therefore be represented easily on maps and map-like objects. Simply mapping spatial data can shed so much more light on what is being studied than if the data are presented in tabular form.³ However, maps can also deceive (see Monmonier [1991] for a popular treatment of this topic) and there are many GI Science issues that need to be considered if spatial data are to be displayed to provide reasonably accurate information content. The development of algorithms for continuous cartograms, software to create pseudo-3D virtual reality environments, and hardware that allows digital video to be linked to a GPS is providing us with the means towards much more sophisticated visualization of spatial data than traditional 2D maps and there are great advances in this area yet to come. Because spatial data contain attribute and locational information, the data can be shown together as on a simple map of the distribution of an attribute or they can be displayed separately in different windows. The use of multiple windows for displaying spatial data is now commonplace and it can sometimes provide useful information to display a map in one window and a non-spatial display of the data (such as a histogram or scatterplot for example) in another window and to provide a link between the two (see GeoDA™ [Anselin, Syabri, and Kho 2006] and STARS [Rey and Janikas 2006] for examples of such systems). In this way, only data highlighted on the non-spatial display need to be mapped to show the spatial distribution of extreme values, for example. Alternatively, all the data can be displayed on a map and only the data points selected on the map need to be highlighted in the non-spatial display. Finally, many spatial data sets are multivariate and it is a major challenge to try to represent such complex data in one display.

The other set of methods to turn spatial data into information are those involving statistical analysis and mathematical modeling. Statistical analysis was traditionally dominated by what is known as “confirmatory” analysis in which a major objective was to examine hypotheses about relationships that were already formed. The typical approach to confirmatory analysis would be to develop a hypothesis about a relationship from experience or the existing literature and to use statistical techniques to examine whether the data support this hypothesis or not. Confirmatory statistical analysis generally depends on assessing the probability or likelihood that a relationship or pattern could have arisen by chance. If this probability or likelihood is very low, then other causes may be sought. The assessment of the role of chance necessitates the calculation of the uncertainty of the results found in a set of data (if we had a different data set, would the results perhaps be substantially different or pretty much the same?). In classical statistical methods, this calculation typically

assumes that the data values are independent of each other. A major problem arises in the use of this assumption in spatial data analysis because spatial data are typically not independent of each other. Consequently, specialized statistical techniques have been developed specifically for use with spatial data (see Bailey and Gattrell [1995] for an informative and accessible summary of some of these techniques) and a great deal more research is needed in this area.

More recently, and probably related to the recent explosion of data availability, “exploratory” statistical techniques have increased in popularity. With these, the emphasis is more on developing hypotheses from the data rather than on testing hypotheses. That is, the data are manipulated in various ways, often resulting in a visualization of the data, so that possible relationships between variables may be revealed or exceptions to general trends can be displayed to highlight an area or areas where relationships appear to be substantially different from those in the remainder of the study region. A whole set of localized statistical techniques has been developed to examine such issues (for example, Fotheringham, Brunson, and Charlton 2000).

Finally, spatial modeling involves specifying relationships in a mathematical model that can be used for prediction or to answer various “what if” questions. Classical spatial models include those for modeling the movements of people, goods, or information over space and the runoff of rainwater over a landscape. There is a fuzzy boundary between what might be termed a mathematical model and what might be termed a statistical model. Quite often, models are hybrids where a formulation might be developed mathematically but the model is calibrated statistically. Where models are calibrated statistically from spatial data, one important issue is that it is seldom clear that all, or even most, relationships are stationary over space, usually an assumption made in the application of various modeling techniques. For instance, the application of traditional regression modeling to spatial data assumes that the relationships depicted by the regression model are stationary over space. Hence, the output from a regression model is a single parameter estimate for each relationship in the model. However, it is quite possible that some or all of the relationships in the model vary substantially over space. That is, the same stimulus may not provoke the same response in all parts of the study region for various contextual, administrative or political reasons – people in different areas, for example, might well behave differently. Consequently, specialized statistical techniques such as Geographically Weighted Regression (GWR) have been developed recently to allow for spatially varying relationships to be modeled and displayed (Fotheringham, Brunson, and Charlton 2002).

The final layer of Figure 0.1 represents some of the application areas of geocomputation which gives an indication of why it is such an important and rapidly growing area of study. Spatial data can be found in most areas of study and include many different types of data, such as:

- Geodetic – coordinate reference systems for locating objects in space;
- Elevation – recording heights of objects above mean sea level;
- Bathymetric – recording the depth of water bodies;
- Orthoimagery – georeferenced images of the earth’s surface;
- Hydrography – data on streams, rivers and other water bodies;

- Transportation networks – roads, railways, and canals;
- Communication networks – the transmission of ideas and data across space;
- Cadastral – precise positioning of property boundaries;
- Utilities – the locations of pipes, wires and access points;
- Boundaries – electoral, administrative, school and health districts;
- Medical – the location of incidents of disease and patients with respect to the location of services;
- Crime – the location of police incidents;
- Environmental – habitats, pollution, and the impacts of natural disasters;
- Urban – the location of areas of high priority for social and economic intervention;
- Planning – the spatial impacts of locational decisions;
- Retailing – the location of consumers with respect to the location of services;
- Biogeography – the location of one species with respect to the location of one or more others.

We now turn to a brief discussion of the topics covered in this book. Given the enormous breadth of GI Science, it is clear that not everything can be included in this volume. However, in order to be as comprehensive as possible, we have tried to solicit contributions which have a fairly general application as opposed to being strictly about the use of GI Science in one particular field.

What Follows Next!

The remainder of this book is organized under six headings. We start each section with a brief description of the chapters that follow and the chapters, themselves, offer stand-alone treatments that can be read in any order the reader chooses. Each chapter includes links to other chapters and key references in case the reader wants to follow up specific themes in more detail.

The first group of six chapters looks at some of the recent trends and issues concerned with geographic data acquisition and distribution. Separate chapters describe how the production and distribution of geographic data has changed since the mid-1970s, the principal sources of social data for GIS, remote sensing sources and data, the possibilities of using spatial metaphors to represent data that may not be inherently spatial for knowledge discovery in massive, complex, multi-disciplinary databases, the myriad sources of uncertainty in GIS, and the assessment of spatial data quality.

The second section of the book explores some of the important and enduring database issues and trends. Separate chapters describe relational, object-oriented and object-relational database management systems, the generation of regular grid digital elevation models from a variety of data sources, the importance of time and some of the conceptual advances that are needed to add time to GIS databases, and new opportunities for the extraction and integration of geospatial and related online data sources.

The third section of the book consists of seven chapters that examine some of the recent accomplishments and outstanding challenges concerned with the visualization of spatial data. Separate chapters describe the role of cartography and interactive

multimedia map production, the role of generalization and scale in a digital world, the opportunities to display and analyze a variety of geographical phenomena as surfaces, fuzzy classification and mapping in GIS, predictive rule-based mapping, multivariate visualization, and the ways in which digital representations of two-dimensional space can be enriched and augmented through interactivity with users in the third dimension and beyond.

The fourth section of the book contains three chapters looking at the increasingly important task of knowledge elicitation. These chapters examine the role of inference and the difficulties of applying these ideas to spatial processes along with the process of geographic knowledge discovery (GKD) and one of its central components, geographic data mining, and the prospects for building the geospatial semantic web.

The next group of four chapters examines spatial analysis. The links between quantitative analysis and GIS, spatial cluster analysis, terrain analysis, and dynamic GIS are discussed here.

The six chapters of Part VI examine a series of broader issues that influence the development, conduct, and impacts of geographic information technologies. Separate chapters examine institutional GIS and GI partnering, public participation GIS, GIS and participatory decision-making, several participatory mapping projects from Central America to illustrate the dynamic interplay between conceptions of people and place and the methods used to survey them, the relationship between GIS, personal privacy, and the law across a variety of jurisdictions, and the major developments and opportunities for educating oneself in GIS.

Finally, Part VII examines future trends and challenges. Separate chapters examine the role of the World Wide Web in moving GIS out from their organization- and project-based roles to meet people's personal needs for geographic information, the emergence of location-based services (LBS) as an important new application of GIS, and two views of the challenges and issues that are likely to guide GI Science research for the next decade or more.

Closing Comments

This handbook seeks to identify and describe some of the ways in which the rapidly increasing volumes of geographic information might be turned into useful information. The brief introductions to topics offered in the previous section give some clues as to what we think is important here – the rapid growth in the number and variety of geographic data sets, finding new ways to store, process, and transmit these data sets, new forms of visualization and statistical/mathematical modeling, etc. To the extent that this book has helped to clarify the current state of knowledge and indicate profitable avenues for future research, it will have helped to educate and inform the next generation of geographic information scientists and practitioners. This generation will need to be more nimble than its predecessors given the rapid rate of technological change (innovation) and the tremendous growth of geographic information science, geographic information systems, and geographic information services that is anticipated in the years ahead. With this in mind, we hope the reader will tackle the remainder of the book with an opportunistic and forward-looking view of the world around them.

ENDNOTES

- 1 In some circumstances, Geomatics also is used synonymously with GI Science as in Geomatics for Informed Decisions (GEOIDE), the Canadian Network Centre of Excellence headquartered at Laval University (www.geoide.ulaval.ca). However, in other circumstances, such as in the naming of academic departments in the UK, the term Geomatics has been used to “re-brand” Departments of Surveying where its scope and purpose are much more restricted.
- 2 While we are used to seeing orthophotographs, photographs with associated files giving information on the location of each pixel so that operations can be carried out on the spatial relationships within the photograph, spatially encoded digital video allows the user to perform spatial queries and spatial analysis on video images. As one can imagine, the volumes of such data that need to be stored and transmitted create special challenges.
- 3 See Ian McHarg’s 1969 book entitled “Design with Nature” for an influential book that documented how maps could be overlaid and used to evaluate the social and environmental costs of land use change. This book has been reprinted many times and still serves as an important text in many landscape architecture courses and programs.

REFERENCES

- Anselin, L., Syabri, I., and Kho, Y. 2006. GeoDa: An Introduction to Spatial Data Analysis. *Geographical Analysis* 38: 5–22.
- Bailey, T. C. and Gattrell, A. C. 1995. *Interactive Spatial Data Analysis*. New York: John Wiley and Sons.
- Burrough, P. A. 1986. *Principles of Geographical Information Systems for Land Resources Assessment*. New York: Oxford University Press.
- Chrisman, N. R. 1999. What does “GIS” mean? *Transactions in GIS* 3: 175–86.
- Corbett, J. D. and Carter, S. E. 1996. Using GIS to enhance agricultural planning: The example of inter-seasonal rainfall variability in Zimbabwe. *Transactions in GIS* 1: 207–18.
- Cowen, D. J. 1988. GIS versus CAD versus DBMS: What are the differences? *Photogrammetric Engineering and Remote Sensing* 54: 1551–4.
- Custer, S. G., Farnes, P., Wilson, J. P., and Snyder, R. D. 1996. A comparison of hand- and spline-drawn precipitation maps for mountainous Montana. *Water Resources Bulletin* 32: 393–405.
- Dykes, J. A., MacEachren, A. M., and Kraak, M. J. (eds). 2005. *Exploring Geovisualization*. Amsterdam: Elsevier.
- Foresman, T. W. (ed.). 1998. *History of Geographic Information Systems: Perspectives from the Pioneers*. Englewood Cliffs, NJ: Prentice Hall.
- Fotheringham, A. S., Brunson, C., and Charlton, M. E. 2000. *Quantitative Geography: Perspectives on Spatial Data Analysis*. Thousand Oaks, CA: Sage Publishers.
- Fotheringham, A. S., Brunson, C., and Charlton, M. E. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester: John Wiley and Sons.
- McHarg, I. L. [1969] 1994. *Design with Nature* (25th anniversary edn). New York: John Wiley and Sons.
- Monmonier, M. 1991. *How to Lie with Maps*. Chicago: Chicago University Press.
- Rey, S. J. and Janikas, M. V. 2006. STARS: Space-time analysis of regional systems. *Geographical Analysis* 38: 67–86.

Part I Data Issues

This first group of chapters looks at some of the recent trends and issues concerned with geographic data acquisition and distribution. The first of these chapters, by David J. Cowen, describes how the production and distribution of geographic data has changed in the past three decades. This chapter walks the reader through the overlapping worlds of the National Spatial Data Infrastructure, Federal Geographic Data Committee (FGDC), framework data, metadata, standards, FGDC clearinghouses, Geospatial One-Stop, and the National Map, and thereby offers a summary of recent developments in the United States where publicly funded geographic data sets have been distributed at little or no cost to potential users for many years.

The second chapter in this group, by David J. Martin, reviews the principal sources of social data for Geographic Information Systems (GIS). The examples demonstrate how conventional administrative, survey, and census-based data sources are becoming increasingly integrated as national statistical organizations look towards data collection strategies that combine elements of each. This integration will probably produce higher spatial and temporal resolution and cross-scale data sets in future years that will, in turn, require entirely new geocomputational tools for effective visualization and/or analysis.

The third of the chapters, by Brian G. Lees, describes the remote sensing sources and data that are commonly used as inputs to GIS. The opportunities for extracting and updating spatial and attribute information in geographic databases from remotely sensed data are examined in some detail, and special attention is paid to the role of error within remote sensing and how insights about the behavior of spatial data in GIS and spatial statistics are feeding back into remote sensing and driving innovation in these rapidly evolving fields.

In the fourth chapter André Skupin and Sara I. Fabrikant explore the possibilities of using spatial metaphors to represent data that may not be inherently spatial for knowledge discovery in massive, complex, multi-disciplinary databases. This area of research is termed “spatialization” and the chapter discusses what kinds of data can be used for spatialization and how spatialization can be achieved. The authors conclude their chapter by noting that spatialization is a new and exciting area in

which GI Science is challenged to address important cognitive and computational issues when dealing with both geographic and non-geographic data.

Ashley Morris starts out the fifth chapter by noting that uncertainty permeates every aspect of spatial data (including the assimilation and storage of geospatial features, operations on those features, and the representation of the results of these operations) and goes on to explain why fuzzy object-oriented databases provide a viable and attractive option for modeling uncertainty in spatial databases. These databases provide membership functions that aid in the storage and representation of objects with uncertain boundaries and the focus on features means that they are able to store and represent both vector- and raster-based objects. These features coupled with the use of multiple alpha-cuts provide extensible systems that can support objects with either crisp or ill-defined boundaries at any desired level of detail. The chapter concludes by noting that the storage, manipulation, and representation of objects with uncertain boundaries is likely to become more important as users become more sophisticated.

The final chapter in this group of six, by James D. Brown and Gerald B. M. Heuvelink, focuses on geographic data, it complements the previous chapter and deals with the assessment of spatial data quality. This information is essential if we are to manage social and environmental systems effectively and more generally, for encouraging responsible use of spatial data where knowledge is limited and priorities are varied. This chapter offers an overview of data quality and measures of data quality, the sources of uncertainty in spatial data, and some probabilistic methods for quantifying the uncertainties in spatial attributes. The conclusion notes several challenges that must be overcome if we are to estimate and use information on spatial data quality more effectively in the future.